RL-TR-96-153
Final Technical Report
September 1996

SPECKLE GENERATION BY PHASE-ONLY SPATIAL LIGHT MODULATORS: Random Phase Properties that Produce Fully Complex Modulation and that Model Optical Processor Performance

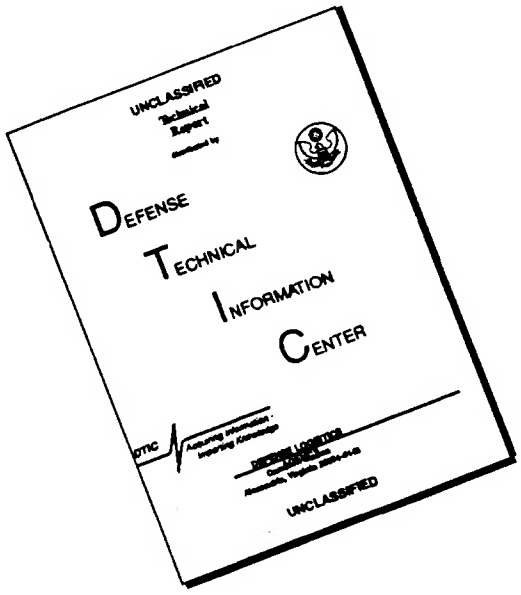
University of Louisville

Sponsored by
Advanced Research Projects Agency

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

Rome Laboratory
Air Force Materiel Command
Rome, New York

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RL-TR-96-153 has been reviewed and is approved for publication.

APPROVED:

THOMAS J. GRYCEWICZ, Major, USAF
Project Engineer

FOR THE COMMANDER:

ROBERT V. McGAHAN
Director
Electromagnetics & Reliability Directorate

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify RL (EROP) Hanscom AFB MA 01731. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

SPECKLE GENERATION BY PHASE-ONLY SPATIAL LIGHT MODULATORS: Random Phase Properties that Produce Fully Complex Modulation and that Model Optical Processor Performance

Robert W. Cohn

Contractor: University of Louisville
Contract Number: F19628-92-K-0021
Effective Date of Contract: 25 June 1992
Contract Expiration Date: 25 March 1995
Short Title of Work: Speckle Generation by Phase-Only Light Modulators
Period of Work Covered: Jun 92 - Mar 96

Principal Investigator: Robert W. Cohn
Phone: (502) 852-7077

RL Project Engineer: Thomas J. Grycewicz, Major, USAF
Phone: (617) 377-3806

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by Thomas J. Grycewicz, Major, USAF, RL/EROP, 80 Scott Drive, Hanscom AFB MA 01731-2909.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1996	Final Jun 92 - Mar 96

4. TITLE AND SUBTITLE
SPECKLE GENERATION BY PHASE-ONLY SPATIAL LIGHT MODULATORS: Random Phase Properties that Produce Fully Complex Modulation and that Model Optical Processor Performance

5. FUNDING NUMBERS
C - F19628-92-K-0021
PE - 61101E
PR - H536
TA - 00
WU - 03

6. AUTHOR(S)
Robert W. Cohn

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Louisville
Department of Electrical Engineering
Louisville KY 40292

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Advanced Research Projects Agency
3701 North Fairfax Dr
Arlington VA 22203-1714

Rome Laboratory/EROP
80 Scott Dr
Hanscom AFB MA 01731-2909

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
RL-TR-90-153

11. SUPPLEMENTARY NOTES
Rome Laboratory Project Engineer: Major Thomas J. Grycewicz/EROP/(617)337-3806

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
Optoelectronic processors that use phase-only and amplitude-phase coupled spatial light modulators (SLMs) can be profoundly affected by phase errors. These errors arise not only from irregularities in the modulators themselves, but at a number of points along the signal processing chain. Traditional theories of speckle generation by rough surface scattering are adapted to analyzing SLMs. SLMs are modeled as arrays of subapertures/pixels that are perturbed by random phase components (constant and linear retardations). The effects of not only random, but also systematic, phase errors on the performance of diffractive optical elements (DOEs) and optical correlators are evaluated. While traditional speckle theory models random surfaces as stationary random processes, SLMs can be programmed to produce nonstationary optical surfaces. This generalization is used to devise a new class of computer generated holographic algorithms, referred to as pseudorandom encoding. The method is notable in that it: 1) uses all available space bandwidth of the SLM; 2) produces diffraction patterns having large signal to noise ratio; and, most notably, 3) can be calculated in real-time by serial processors. Simulations and experimental demonstrations using SLMs are presented, and applications to pattern recognition, optical interconnects, multi-spot beam steering, and acceleration of DOE design algorithms are described.

14. SUBJECT TERMS Optical information processing, Phase-only spatial light modulators, Composite pattern recognition filters, Optical correlators, Laser speckle, Rough surface scattering, Statistical optics, Correlation metrics, Multi-spot phased array beam steering

15. NUMBER OF PAGES 166

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

SPECKLE GENERATION BY PHASE-ONLY SPATIAL LIGHT MODULATORS:
Random Phase Properties that Produce Fully Complex Modulation
and that Model Optical Processor Performance

Table of Contents

SPECKLE GENERATION BY PHASE-ONLY SPATIAL LIGHT MODULATORS:
Random Phase Properties that Produce Fully Complex Modulation and that Model Optical Processor Performance

ABSTRACT

Optoelectronic processors that use phase-only and amplitude-phase coupled spatial light modulators (SLMs) can be profoundly affected by phase errors. These errors arise not only from irregularities in the modulators themselves, but at a number of points along the signal processing chain. Traditional theories of speckle generation by rough surface scattering are adapted to analyzing SLMs. SLMs are modeled as arrays of subapertures/pixels that are perturbed by random phase components (constant and linear retardations). The effects of not only random, but also systematic phase errors on the performance of diffractive optical elements (DOEs) and optical correlators are evaluated. While traditional speckle theory models random surfaces as stationary random processes, SLMs can be programmed to produce nonstationary optical surfaces. This generalization is used to devise a new class of computer generated holography algorithms, referred to as pseudorandom encoding. The method is notable in that it 1) uses all available space bandwidth of the SLM; 2) produces diffraction patterns having large signal to noise ratio; and, most notably, 3) can be calculated in real-time by serial processors. Simulations and experimental demonstrations using SLMs are presented, and applications to pattern recognition, optical interconnects, multi-spot beam steering and acceleration of DOE design algorithms are described.

1.0 INTRODUCTION

1.1 Technical problem and objective of the study. Phase errors can profoundly distort the intended diffraction patterns of spatial light modulators (SLMs) thus impacting the performance of optical processing systems that use SLMs. Phase errors are introduced by a variety of mechanisms. Phase-modulating SLMs can have inherent phase errors, due to fabrication process variations from pixel to pixel. Noise (e.g. thermal, quantization, etc.) on the video signals modulating SLMs is also transformed into phase errors. Furthermore, many applications, including composite filters for pattern recognition, binary diffractive optics and optical neural networks, can often be better understood and analyzed by modeling the modulations/signals as random, rather than as deterministic.

Random phase modulations across the surface of the SLM diffract into broadly spread noise patterns. These noise patterns not only have the appearance of speckle patterns, but in fact, arise from the identical situation of scattering of light from a random surface. There is a wealth of information on laser speckle (J.C. Dainty, *Laser Speckle and Related Phenomena*, Springer,

1984) and statistical optics (J. W. Goodman, *Statistical Optics*, Wiley, 1985) that is applicable to SLM-based optical processors. This study was motivated by our opinion that speckle theory could be applied in new ways to advance both performance modeling of optical processors, and to lead to new applications of SLMs. As we report, advancements in both applications and analysis have been made as a result of this approach.

1.2 General methodology. This study included the mathematical modeling and computer simulation of diffraction from phase-only and coupled amplitude-phase SLMs. Models and simulations of the performance of optical correlators that use phase-only SLMs were also derived as specialized cases of the analysis methodology. Experimental confirmation of our theories using real devices was another major emphasis of the study. This provides better appreciation of the theories, allows us to refine the models based on the outcome of the experiments, and also enhances the art of using SLMs. The combined emphasis on theory and experiment also led to improved SLM measurement/characterization techniques which are reviewed below. The experiments included experimental demonstrations of spot arrays designed to have up to 100 uniform intensity spots, and single spots that were simultaneously steered and shaped from circular to elliptic with up to a 5:1 eccentricity.

During this study there were many discussions on properties of SLMs with SLM developers (including Meadowlark Optics, Hughes, Texas Instruments, Hamamatsu, Boulder Nonlinear), SLM users (Rome Labs Hanscom and Griffiss; U. S. Army Missile Command; Army Research Lab, MD; R. Juday, T. H. Chao, and J. Downie of NASA; Dr. Y. Sheng, U. of Laval; TOPS participants: Teledyne Brown and Martin Marrieta) and diffractive optics groups (P. Maker of NASA JPL and with Teledyne Brown). Results of this study were communicated to these groups and feedback from these groups influenced the focus of the study. An additional benefit of the discussions is that Hughes, subsequent to this study, provided us with custom made light valves having phase ranges in excess of 360°.

1.3 Report outline. This report is written to provide enough background so that even readers who are not specialists in the areas of optical processing, scalar diffraction, rough surface scattering and statistics will be able to evaluate the technical approach and accomplishments of the study. The reports begins by comparing the assumptions of traditional models of speckle generation from rough surfaces with the assumptions needed to model speckle generation by spatial light modulators. We then present a summary of the mathematical models that describe speckle generation. These models are then used to develop performance analyses of optical processors and to develop algorithms that approximate fully complex modulation on SLMs that have a constrained (i.e. phase-only or coupled amplitude-phase) modulation range. The latter procedure we refer to as *pseudorandom encoding*. The most significant simulated and experimental results are also presented.

More detailed information and additional specific results are presented in the attached publications that were written during the study. These are liberally referenced throughout the report and may be consulted by the reader desiring additional detail. Citation numbers appear as superscripts and

refer to the contract-supported publications list in Sec. 7.0.

This outline concludes with a summary of significant findings and a list of recommendations for future studies. To most readers this summary is too brief to follow in the first reading of the report. However, in future readings it can be used as an index to directly guide the reader to specific results of interest in the report and the attached papers.

Summary of Significant Findings

- Closed form expressions for the expected intensity and standard deviation of the Fourier plane diffraction pattern from arrays of independent non-identically distributed random complex-valued pixels were derived and verified for correctness and numerical efficiency.[1,7] The expressions were applied 1) to the analysis of diffraction from SLMs having random phase piston and tilt errors;[1] 2) to the effect of random phase errors on phase-only optical correlators;[2] and 3) to the development of an *in situ* method of measuring the phase modulation depth of SLMs.[9] (Sec. 3.0)

- Mathematical models of the effect of systematic phase errors (including quantization, linear and quadratic gain, and saturation) on the performance of phase-only correlators were developed and verified against simulations.[3,4] (Sec. 5.0)

- The above two analyses were combined into a single model and applied to error analyses of SLMs in optical correlators having combined random and systematic filter plane phase errors.[5,6] (Sec. 5.0)

- A statistically based algorithm *pseudorandom phase-only encoding* was developed which uses pseudorandomly selected values of phase to represent arbitrary complex values on a phase-only SLM.[7] (Sec. 3.1) The method has been demonstrated both with simulations[7,8,11,13-15] and experiments[9,13-15] using SLMs to produce highly accurate diffraction patterns having low levels of noise. Specific designs have been demonstrated that include beam steering and shaping,[7-9] (Fig. 4), multiple spot beam steering,[13-15] (Sec. 4.2) and composite/synthetic discriminant function filters for pattern recognition.[10,11] (Sec. 4.5) Error analyses show that the quality of the diffraction patterns is closely related to the diffraction efficiency η and it can be easily calculated prior to actually performing the encoding.[7,9,11] (Sec. 3.2) The encoding method, as opposed to most numerically intensive diffractive optic design procedures, can be performed in real-time, thus making it well suited for real-time and adaptive optical processors.

- Pseudorandom encoding was extended in several ways including 1) improving performance by only selective or partial encoding of the desired complex values,[11,15] (Secs. 4.3,4.4); 2) developing encoding formulae for modulators for which the value of amplitude is a function of the value of phase,[16,17] (Sec. 3.1); 3) inventing a type of fixed-

pattern diffractive optic, referred to as a *patterned diffuser array* together with a low cost, high speed fabrication method.[12,13] (Secs. 4.6-4.9) These diffractive optical elements (DOEs) effectively produce fully-complex modulation and have much lower noise than possible for the encoding method for SLMs. It was also found that partial pseudorandom encoding produced higher performance composite pattern recognition filters on phase only SLMs than Horner's phase-only mapping, Juday's minimum euclidean distance (MEDOF) design procedure, or pseudorandom encoding by itself.[11]

- Detailed characterizations of the phase modulating properties of the TVT6000 liquid crystal SLMs were performed including characterizations of the effects of modulation transfer function (MTF) on phase depth.[9]

Recommendations for Future Studies are to

- Extend analyses of random errors to Fresnel diffraction and investigate the possibility of pseudorandom synthesis in this regime.

- Extend models of systematic and random errors of correlators to specifically analyze 1) coupled amplitude phase SLMs; 2) composite/SDF filters.

- Develop an analytic procedure for determining the best operating curve from a range of curves (such as the polarization dependent properties of liquid crystal SLMs) so that a correlator optimally recognizes the target. The goal is to develop a method of selecting this curve that can be solved in real time.

- Develop a pseudorandom encoding formula for an available (i.e. coupled) SLM and experimentally demonstrate pseudorandom encoding with this SLM. Experimentally demonstrate and characterize the performance of optical correlation using encoded SLMs.

- Investigate ways that pseudorandom encoding could be used to accelerate current diffractive optic and composite filter design methods over the current numerically intensive solution methods.

- Demonstrate a prototype system to fabricate patterned diffuser arrays.

- Demonstrate usefulness of patterned diffuser arrays as accurate, easily fabricated grayscale masks in projection printers for photolithographically patterning three dimensional surfaces for diffractive optics, micro-optics and other applications requiring topographic structures.

- Further evaluate usefulness of pseudorandom grating in distributed feedback lasers and couplers, acousto-optic tunable filters, artificial dielectric/submicron scale gratings, and for thick and volume holographic recording.

2.0 QUALITATIVE DESCRIPTION OF SPECKLE GENERATION

2.1 Basic concepts of rough surface scattering and generation of speckle. *Speckle patterns* are produced by scattering plane waves off of rough surfaces and observing the resulting pattern of intensity at some distance from the surface. This is illustrated in Fig. 1. Surfaces that are perfectly smooth produce mirror reflection. The resulting far field pattern is referred to as the *specular component*. The diffraction pattern from a very rough surface (for which the rms optical path differences are much greater than a wavelength) instead produces diffuse scatter which is observed as a speckle pattern. The high spatial frequency of the surface causes the light to diffract over a much greater angular range than the specular component, and the (typically) random texture of the roughness gives the speckle pattern its noise-like intensity pattern. Fig. 1 only illustrates the average intensity envelope of the speckle pattern. The shape and spread of the envelope is due mainly to the size of the individual rough grains. Surfaces that are less rough produce *partially developed speckle patterns* which are composed of the broadly diffused speckle pattern, plus a specular component of reflection. Fig. 1 illustrates that it is possible to vary the intensity of specular light by controlling roughness.

The formation of speckle can be viewed as the random phasing of plane wavefronts at a great distance from the surface. This is specifically illustrated in Fig. 2. The wavefronts arise from an array of N point sources (which might model a surface or a spatial light modulator). We will consider the effect of different types of source distributions on the far field pattern. If all the wavefronts are of the same phase then they will reinforce each other and produce an intense light distribution on the optical axis. This corresponds to the purely specular diffraction pattern in Fig. 1. If the phases of the point sources are randomly distributed between 0 and 2π radians then the wavefronts interfere with all different phases and the observed intensity is much weaker. This corresponds to the purely diffuse diffraction pattern in Fig. 1. The percentage of diffuse to specular energy can be varied from 0 % to 100 % by increasing the randomness of the phases from 0 to 2π.

The superposition of randomly phased wavefronts (as illustrated in Fig. 2) can also be interpreted as performing statistical averaging. The concept is similar to Monte Carlo simulation methods. If a large number of repeated experiments or trials are performed and the results are averaged together then the average result is approached with increasing accuracy as the number of trials is increased. This is often referred to as *the law of large numbers*. For example, considering simply pouring a stream of salt onto the ground. After only a few grains are scattered, the distribution of grains appears random, but over a period of time a mound forms that becomes increasingly well defined, smooth and deterministic. From this statistical viewpoint, the superposition of wavefronts is identical in concept. Increasing the number of interfering wavefronts results in diffraction patterns that more accurately approach the average (i.e. the expected value) of the diffraction pattern. This study is heavily based on this particular viewpoint. Since SLMs have a moderately large number of pixels (10,000 to one million,) we have observed and report diffraction patterns that very closely approximate their true average (that would result for an infinite number of pixels.)

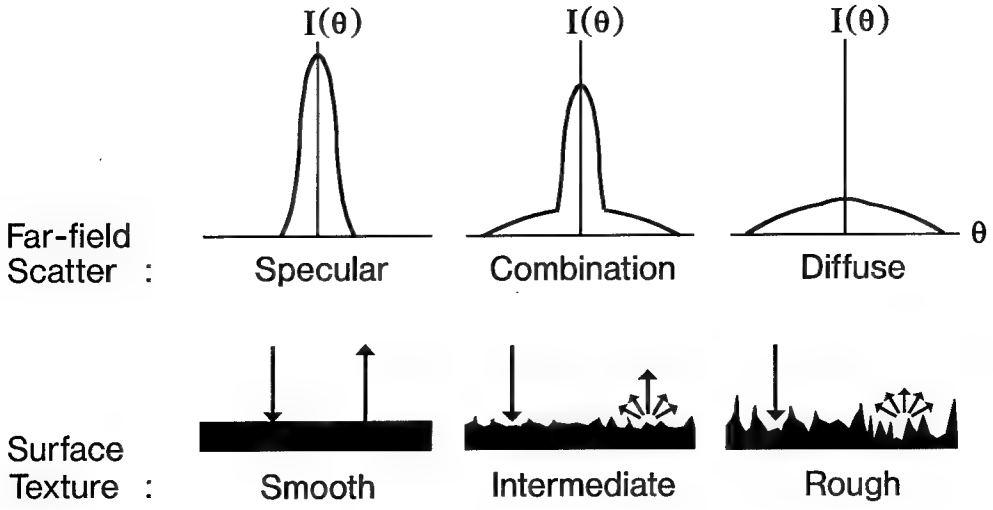


Fig. 1. Rough/random phase used to control specular intensity

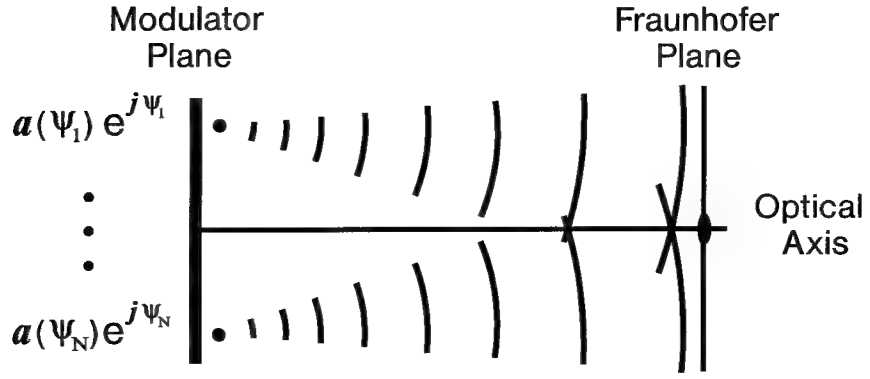


Fig. 2. Typical phased array geometry for which large numbers of wavefronts from SLM are added together in far-field.

Extension of these basic concepts leads to (1) new methods of evaluating the performance of optical processors and (2) new and faster methods of programming fully complex-valued (i.e. arbitrary amplitude and phase) spatial modulation onto limited range (e.g. phase-only) SLMs. We refer to these new design algorithms as *pseudorandom encoding*. Mathematically, it can be viewed as design based on statistical averages. Physically, it can be interpreted as the result of light scattering from rough surfaces. These surfaces are more general than traditional speckle theories, especially in that the statistical properties of each SLM pixel can be selected to *effectively* (i.e. on average) produce a desired modulation.

2.2 Comparison of traditional speckle models with models needed for SLMs. A significant amount of research has been devoted to statistical modeling and experimental characterization of speckle properties (see Dainty, cited above.) These models have focused on naturally rough types of surfaces such as machined surfaces, paper, and biological tissues. The surface is typically

modeled as a plane perturbed by a roughness function that is a stationary random process (i.e. the statistics of roughness are unchanging with position across the surface.)

The most fundamental difference between the traditional surface models used in the theory of speckle and the models developed in this study are that we generalize the surface so that it is modeled by a *nonstationary* process. This is to say that the statistics of the process (thus the texture of the roughness) varies with position. The nonstationary model is required to synthesize diffraction patterns by the pseudorandom encoding method (Sec. 3.1) in which the randomness (corresponding to roughness) in a particular position on the surface determines the amount of light specularly reflected from that position (see Fig. 1.)

A second difference between traditional surfaces and surfaces possible with SLMs becomes apparent when considering SLMs that are composed of arrays of individually programmable pixels (or equivalently, when the signal used to program the SLM is discrete rather than continuous.) SLM pixels have finite width apertures that can modify the sample values of the signal. This transformation, referred to as a sampling effect, can modify the signal in simple or complicated ways, depending on the exact modulation properties of a pixel. A pixel may simply convert a discrete signal value into a constant value of modulation across its aperture (akin to a sample and hold circuit), or it may produce a varying modulation across its aperture [for example, a cantilever beam Deformable Mirror Device (DMD) which produces a linear phase ramp.] While, natural surfaces typically lack these deterministic structural details, we have found it necessary to include pixel structure in our SLM models.

A similarity between our models of the signal and traditional models of the roughness function is that each are usually assumed to be composed of statistically independent samples. This is equivalent to saying that the function is statistically uncorrelated with shifted versions of itself. However, once the signal is applied to program the SLM, some spatial correlation can be introduced. For instance, pixels that convert signal samples into a constant value over the pixel aperture cause the SLM surface to be correlated over that the width of the aperture. Limited spatial resolution of an SLM can also lead to correlated spatial modulation. For the experimental studies that we performed using liquid crystal televisions (LCTVs) we found that the electronic drive circuitry actually lowpass filtered the video signal causing the correlation distance to be greater than the pixel spacing.[9] Thus the resolution of these devices is actually significantly lower than the spatial frequency of the pixels. In experimental comparisons with theories assuming uncorrelated samples, we took steps to minimize the degree of correlation; most notably, programming clusters of pixels identically as a superpixel.

In this study we specifically modeled SLMs as arrays of pixels that are deterministic, but which can be perturbed by random parameters, e.g tilt (phase ramp) and piston (phase retardation).[1,2,7] In addition, the random parameters are statistically independent and (in the most general case) non-indentically distributed (which is comparable to a white but non-stationary stochastic process.) Early in the program SLMs with pixels that produce both piston and tilt (representative of deformable mirror devices) were modeled. When it became apparent that Texas Instruments

would not be able to complete fabrication of Flexure Beam DMDs (FBDMD), the modeling focused in on devices with ideal phase-only (piston) modulating pixels. Towards the end of the study we recognized that the analyses developed could be applied with some modification to modulators for which amplitude is a function of phase--- the so called *amplitude-phase coupled modulator*. An example of a coupled modulator characteristic (illustrated both in rectangular and polar) is given in Fig. 3. Most current SLMs have coupled amplitude phase modulation. Most notable are liquid crystal SLMs which can be varied from amplitude-mostly to phase-mostly modulation depending on the polarization of the illumination.

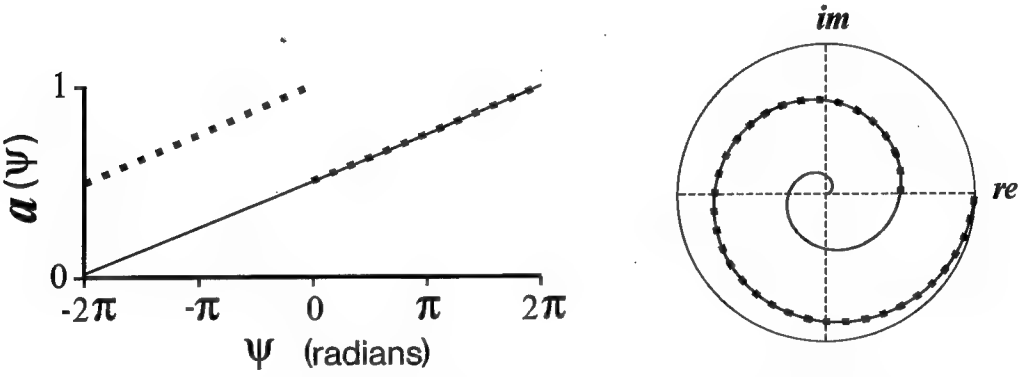


Fig. 3. Amplitude coupling compared for identical slopes with and without phase modding. Shown both as cartesian and polar functions.

3.0 MATHEMATICAL MODELS OF SPECKLE GENERATION.

In this section we mathematically describe the two properties of scattering from randomly modulated SLMs that permitted us to analyze the performance of optical processors (Sec. 4.0) and that we used to develop pseudorandom encoding algorithms. (Sec. 5.0) The first property (referred to as the *complex modulation property*) is that an average, or *effective*, complex value of modulation can be ascribed to randomly modulated SLM pixels for purposes of modeling the far-field diffraction pattern. The second property (referred to as the *noise* or *speckle background property*) is that the far-field pattern will contain a noise/speckle background in proportion to the randomness of the SLM modulation pattern. The fundamentals of these properties are presented in this section. Then in Secs. 4.0 and 5.0 detailed applications of the complex modulation property to designing diffraction patterns and of the noise/speckle background property to evaluating the performance of SLM-based optical processors are presented.

3.1 Complex modulation property.[7] This property is most easily shown for phase-only surfaces. A plane wave reflected from a phase-only surface can be represented by the complex-valued (indicated by bold) function $\boldsymbol{a}(x,y) = \exp[j\psi(x,y)]$. The phase modulation $\psi(x,y)$ is a random field consisting of independent non-identically distributed random variables with co-ordinates x and y. The statistics of the phase are then fully specified by the probability density function (pdf)

of phase $p[\psi(x,y)]$ which varies with position. These assumptions about SLMs correspond to those given in the previous section. With these definitions we can directly calculate 1) the complex amplitude of the far-field diffraction pattern; 2) the expected complex amplitude of the far-field diffraction pattern; 3) the expected intensity of the far-field diffraction pattern; and 4) the standard deviation of the diffraction pattern intensity.

The far-field diffraction pattern of any modulation pattern $\boldsymbol{a}$ is known to be

$$\boldsymbol{A}(f_x, f_y) = \mathscr{F}[\boldsymbol{a}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \boldsymbol{a}(x,y)\exp\left[j2\pi(f_x x + f_y y)\right] dx\,dy \qquad (1)$$

where $\mathscr{F}[\bullet]$ is the fourier transform operator. The expectation or ensemble average of any random variable ψ is written

$$\langle\psi\rangle = \int_{-\infty}^{\infty} \psi\, p(\psi)\, d\psi \qquad (2)$$

where $\langle\bullet\rangle$ denotes the expectation operator. The average complex-amplitude of the far-field pattern of the random complex modulation $\boldsymbol{a}$ can be written

$$\langle\boldsymbol{A}\rangle = \mathscr{F}[\langle\boldsymbol{a}\rangle] \qquad (3)$$

where the linearity of the fourier transform and ensemble average operators has been used to interchange their order. Under the assumption that the random samples of $\boldsymbol{a}$ are statistically independent with position, the expectation of I the far-field intensity pattern is

$$\langle I\rangle = \langle|\boldsymbol{A}|^2\rangle = |\langle\boldsymbol{A}\rangle|^2 + \langle I_s\rangle \qquad (4)$$

where $I_s(f_x,f_y)$ is a residual noise pattern due to the random phasings in the far-field. As long as the noise [represented by the second term of eq. (4)] is adequately low, then eq. (4) is approximately the magnitude squared of eq. (3). In this average sense, any complex-valued modulation can be represented by the random phase-only modulation $\boldsymbol{a}(x,y) = \exp[j\psi(x,y)]$. This is seen by explicitly taking the ensemble average of $\boldsymbol{a}$ which yields

$$\langle\boldsymbol{a}\rangle = \int p(\psi)\exp(j\psi)\, d\psi \equiv a_p \exp(j\phi_p) \qquad (5)$$

where a_p is the resulting expected amplitude modulation. We will often refer to a_p as the *effective amplitude*, $\phi_p = \langle\psi\rangle$ as the *effective phase*, and $\boldsymbol{a}_p \equiv \langle\boldsymbol{a}\rangle$ as the *effective complex amplitude* or *modulation*.

Eq. (5) demonstrates the first property: that randomness can be used to effectively produce fully complex modulation. Even though the actual (non-averaged) modulation is phase-only modulation, each pixel can represent, in an average sense, any desired amplitude $a_p(x,y)$ between 1 and 0. To show this amplitude control explicitly, consider the family of uniform density functions having random phase spreads $v \in [0,2\pi]$. Evaluating the uniform family in eq. (5) gives all values of amplitude between 0 and 1 according to

$$a_p = \mathrm{sinc}(\nu / 2\pi) \quad . \tag{6}$$

Eq. (6) shows that it is possible to select a desired effective amplitude by choosing an appropriate value of ν. The value of ν can be different for each pixel (non-identical/non-stationary statistics) thus enabling near arbitrary diffraction patterns. This observation is the basis for us proposing *phase-only pseudorandom encoding*.

The encoding algorithm for an array of N pixels having the desired complex amplitudes $a_{c\,i}$ is:

for $i=1$ to N:

a) invert eq. (6) for ν_i given the value of $a_{p\,i} = |a_{c\,i}|$

b) select a number between -½ and ½ using a uniform random number generator

c) scale this number by ν_i and offset by $\phi_{p\,i}$ to get the pseudorandom phase ψ_i

Specific diffractive optic designs produced using this algorithm will be reviewed in Sec. 4.0. Before proceeding to the second property of diffraction it is also worth considering how pseudorandom encoding for phase-only SLMs can be extended to coupled amplitude-phase SLMs.

Complex modulation property for coupled amplitude phase SLMs.[16,17] Most practical and currently available SLMs are actually coupled. The current model can be extended so that rather than the amplitude a being unity, its value $a(\psi)$ is now a deterministic function of phase ψ. Under this assumption eq. (5) is generalized to

$$\langle a(\psi)\exp(j\psi)\rangle \equiv \int_{-\infty}^{\infty} a(\psi)\exp(j\psi)p(\psi)\,d\psi \quad . \tag{7}$$

The phase coupled amplitude $a(\psi)$ can be interpreted as adding a weighting function to the phasor average of eq. (5). Our approach to developing a pseudorandom encoding algorithm for coupled modulators was to make eq. (7) look like the effective amplitude for phase-only pseudorandom encoding in eq. (5). This can be done by specifying an effective probability density function of the form

$$p_{eff}(\psi) \equiv a(\psi)p(\psi) \tag{8}$$

so that the phase pdf $p(\psi)$ compensates for the amplitude weighting/coupling $a(\psi)$. For instance, one can select the product of the effective amplitude and the phase pdf to equal a uniform function. Then eq. (7) can be integrated to give a sinc function similar in form to eq. (6).

Note that individual non-uniform random number generators are needed depending on the particular coupling between amplitude and phase. One direct way to synthesize random numbers with the required pdf $p(\psi)$is to transform a uniform distribution into the required pdf according to

$$\psi = P^{-1}(s) \tag{9}$$

where s is a uniform random variable contained between 0 and 1, and $P(\psi)$ is the cumulative distribution function of ψ [i.e. $p(\psi) = dP(\psi)/d\psi$]. In the field of image processing, this

procedure for transforming statistics is referred to as *histogram equalization*. With these modifications it is possible to use randomness to overcome the limited modulation range of many types of SLMs in addition to phase-only SLMs.

For either phase-only or coupled SLMs, pseudorandom encoding offers a direct, pixel-by-pixel method of encoding. Since there are only a few numerical calculations per pixel the method has the potential for real-time implementation. This is the major advantage of the algorithm over most current techniques that search for globally optimal solutions, and which usually require hours of computer time. In the 1960s and 1970s fast encoding techniques were developed for computer generated holography. These work by clustering groups of pixels together to achieve an effective complex value. These methods are as fast as pseudorandom encoding but by clustering pixels they reduce the available spatial bandwidth of the SLM. Considering the small number of pixels in current SLMs (100x100 to 1000x1000) as compared to holograms, we feel that the bandwidth loss from clustering is unacceptable. Pseudorandom encoding does use the full bandwidth of the SLM, which we believe is another of its major advantages.

3.2 Noise/speckle background property.[7] The term $\langle I_s \rangle$ in eq. (4) can be interpreted as the average noise or speckle background that is generated by randomness in the modulation. This noise envelope corresponds to the diffuse background in Fig. 1. The other term $|\langle A \rangle|^2$ in eq. (4) corresponds to the desired specular component in Fig. 1. The specular component can be much more complicated than the single diffraction peak drawn, since it results from the arbitrary phasing of a large number of SLM pixels (as illustrated in Fig. 2).

While eq. (4) does provide information on the expected specular diffraction pattern and the expected background noise, it does not (directly) provide information on the variance of a particular design. In concept, the variance can be directly calculated according to

$$\sigma_I^2(f_x, f_y) = \langle I^2 \rangle - \langle I \rangle^2 \qquad \textbf{(10)}$$

where $I \equiv |A|^2$. In practice, the mathematical manipulations can be quite involved. A very general derivation for an array of pixels is presented in Ref. 7. We have found that the resulting expression can greatly simplify the effort of rederiving the variance over performing the operations using the fundamental operations described in eq. (10) for each new SLM characteristic. The results are general enough to apply to both phase-only and coupled amplitude phase SLMs and there is also no constraint on the spatial properties of the pixel. For example, we have used these general equations to evaluate the performance of SLMs using phase-only pixels having random tilts.[1,2] Presentation of the specific equations in this summary would provide little additional insight. Instead, Ref. 7 (contained in Sec. 10) can be consulted as needed.

The most direct application of eq. (10) is in simply calculating the error bounds for the far-field diffraction pattern as produced by a randomly perturbed SLM. An example of this is illustrated in Fig. 4. The curves shown are one dimensional cross sections from a computer simulation of diffraction from a phase-only (piston-only) SLM. The specific pattern was designed (using pseudorandom encoding) to give an equiripple (i.e. Chebychev) approximation to a brick wall

response. The central curve of the three curves drawn with thin lines represents the expected intensity pattern calculated using eq. (3) and the outer curves are the one standard deviation error bars calculated using eq. (10). The error bounds give an idea about the quality of the design. The actual diffraction pattern (thick curve) compares closely with the average result (thin lines). That the fluctuations of the actual diffraction exceed the one standard deviation error bounds is not unexpected. If we had wanted to show a bound on almost all errors we would have instead plotted the 3σ error bounds. Grayscale plots of the diffraction pattern are also presented in Ref. 7. Comparable results are experimentally demonstrated with an available SLM in Ref. 9.

The equations for expected complex amplitude, expected intensity and standard deviation of diffraction patterns [eqs. (3),(4) and (10)] have also led to the development of analytic expressions for important performance metrics, including signal-to-noise ratio (SNR), peak-to-noise ratio (PNR) and diffraction efficiency η. The exact meaning of these metrics can vary depending on the specific problem analyzed. Some specific results are presented to give more details on the properties of eqs. (3),(4) and (10) and how these are used to develop metrics.

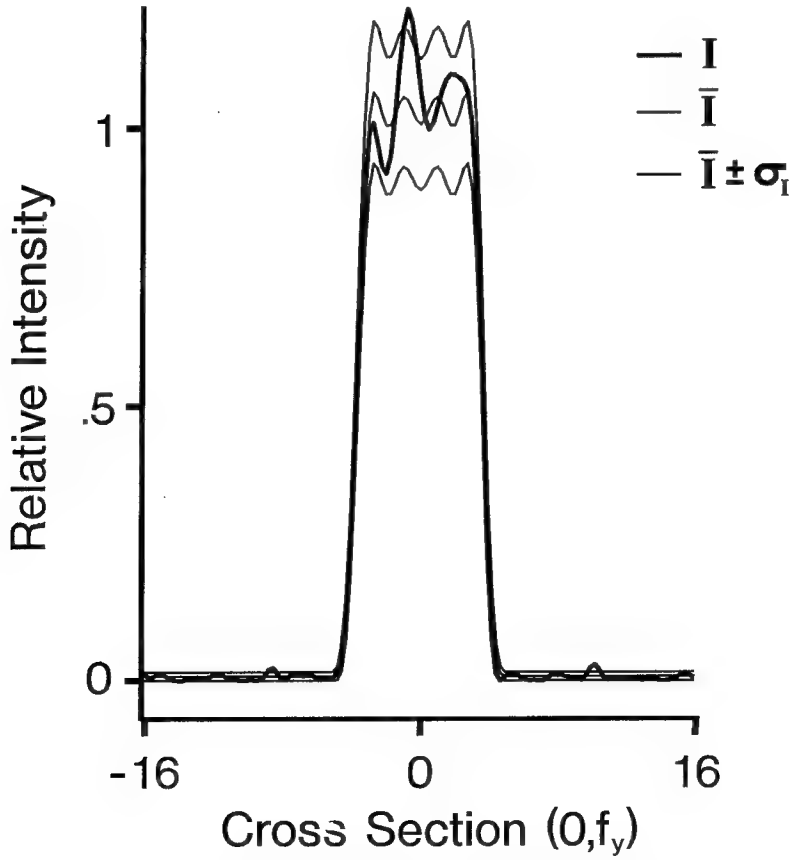


Fig. 4. Brick wall diffraction pattern achieved by pseudorandom encoding.

Metrics that describe the influence of the noise. We will describe the *diffraction efficiency* for pseudorandom encoding of phase-only SLMs. Diffraction efficiency generally measures the ratio of energy that diffracts into the desired/design pattern compared to the total energy available. The definition can vary depending on whether the transmittance loss of the SLM is accounted for. Phase-only SLMs are sometimes stated as being 100% diffraction efficient because they introduce no transmission losses. However this often has little practical relevance because there will be implementation losses due to the limited (i.e phase-only) modulation range of the SLM when designing an arbitrary diffraction pattern. For phase-only pseudorandom encoding we have chosen to define diffraction efficiency η as the ratio of energy that goes into the desired diffraction pattern compared to the total energy (which is the sum of energies from both the desired and the speckle/noise diffraction patterns.) We found this expression for diffraction efficiency by summing up the energy in the corresponding desired and noise terms in eq. (4). The diffraction efficiency for a pseudorandom encoded phase-only (piston type pixels) SLM is then found to be[9,11]

$$\eta = \frac{1}{N} \sum_{i=1}^{N} a_{p\,i}^{2} \tag{11}$$

where N is the number of pixels in the SLM and $a_{p\,i} = |\langle a_i \rangle|$. The ratio of the energy that appears as speckle is then $1-\eta$. In Ref. 7 we originally interpreted the quantity $(1-\eta)N$ as an *effective number of random pixels*. Diffraction efficiency or effective number give a clear physical picture and a mathematical measure of the quality of any pseudorandom encoded design. Eq. (11) is especially important because it is expressed in terms of the desired complex modulation. Thus the quality of each pseudorandom design is known in advance of actually performing the encoding. In Secs. 4.3 and 4.4 we describe how this metric has been used to rapidly search for the highest efficiency diffraction pattern by evaluating the efficiencies of the set of Fourier transform pairs that all produce the same intensity pattern (but which produce different phase patterns).

We can also show to what degree diffraction efficiency influences performance as measured by SNR and PNR metrics. A conventional definition of signal-to-noise ratio is $\mathrm{SNR}=\langle I \rangle/\sigma_I$. We specifically considered single spot diffraction patterns (e.g. shown in Fig. 4 and in Ref. 7) that were designed by phase-only pseudorandom encoding. The SNR metric developed specifically measures the ratio of diffraction peak intensity to its fluctuation at that same point. The reciprocal of SNR is the relative error of the peak intensity of the diffraction pattern. We found[7] that the SNR can be approximated for this case as

$$\mathrm{SNR} \approx \sqrt{\frac{N}{2} \frac{\eta^2}{1-\eta}} \tag{12}$$

where we have expressed the result in terms of diffraction efficiency rather than effective number (as it was originally presented in Ref. 7). The result clearly shows the influence of the number of SLM pixels and of the diffraction efficiency on the accuracy of the resulting diffraction pattern. To gain an appreciation of the relationships between SNR, η, N and the resulting diffraction patterns specifically consider the result illustrated by Fig. 4. For this case the 128x128 pixel SLM has N=16,384 pixels, an SNR of 8:1 and a diffraction efficiency of 8%.

A second measure describing the quality of single spot diffraction patterns is PNR. This can also be described a a measure of peak intensity to standard deviation. The distinction with SNR is that for PNR the standard deviation is measured away from the spot in the noise background. This definition of PNR is often referred to as SNR in many publications (but we use the term PNR to distinguish from our first definition of SNR.) Further evaluation of eqs. (3), (4) and (10) showed that PNR is closely related to SNR and η according to

$$\mathrm{PNR} \approx 2\,\mathrm{SNR}^2 = \frac{N\eta^2}{1-\eta} \quad . \tag{13}$$

Continuing with the numerical example from the previous paragraph we see that the PNR for the example in Fig. 4 is calculated to be 128:1 . These correspond well with the resulting pattern (the thick curve in Fig. 4) in which the individual speckles are barely discernable on the plot.

These results give an idea of the usefulness of the analysis equations [eqs. (3), (4) and (10)] for modeling performance. The specific results described here on the performance of phase-only pseudorandom encoding provide both engineering design criteria and physical insight. The analysis equations are easy to apply to a variety of modeling problems in optical processing. A detailed application of the noise property to modeling the performance limits of optical correlators is presented in Sec. 5.0. We first describe in Sec. 4.0 an application of the complex modulation property.

4.0 COMPLEX MODULATION PROPERTY USED TO DESIGN COMPOSITE FILTERS

This section describes our most advanced procedure, built around pseudorandom encoding, for mapping sophisticated composite functions onto SLMs. Algorithms for encoding fully complex composite functions onto limited modulation range SLMs and that can be performed with a small number of numerical operations in real- or near real time would greatly extend the performance and flexibility of optical processors by making them adaptive to changing and unpredictable situations. For instance, the algorithm to be described could be used to program a phase-only SLM to adaptively and independently steer a number of spots to arbitrary locations in the far-field. Such a system is comparable to a phase-only phased array antenna. The algorithm could also be used to construct in real time composite filters for optical correlators. For instance, the composite filters could be constructed on-the-fly to recognize a particular subset of objects (say tank, humvee and mortar) or a specific subset of views of an object (for distortion invariant recognition and tracking). However, multispot beam steering or distortion invariant pattern recognition is quite difficult to perform if only phase (but not amplitude) of the SLM is controlled. The pseudorandom encoding algorithm described so far overcomes the problems in designing the correct SLM modulation in real-time but it also has a low diffraction efficiency compared to that possible using numerically intensive algorithms (requiring hours of computation.) Our generalized algorithm can produce higher diffraction efficiencies than the original pseudorandom encoding algorithm and the generalized algorithm can run in near-real time (or with further development effort, in real-time.)

4.1 Definition of composite function. Two broad classes of composite functions are the composite pattern recognition filter and the spot array generator. Either class of functions can be viewed as linear combinations of real and/or complex valued functions. The composite filter is used for recognizing an object in the presence of distortions. Being able to identify the same three dimensional object from various two dimensional views is an example of *distortion invariant recognition*.[10,11] Distortion invariance can be built into the filter for a matched filter correlator by adding together filters that recognize the object from its individual views. The *spot array generator* is a diffractive optic that produces an array of equal intensity spots. These have many applications to machine vision and robotic inspection, and to optical processors where they can be used as fan outs for free-space optical interconnects and neural networks. The ideal modulation required for a spot array generator corresponds to a composite function that is a linear combination of complex sinusoids. This would then produce diffraction spots at locations corresponding to the spatial frequency of each sinusoid.

4.2 The problem of mapping composite functions onto limited range SLMs. The problem with any algorithm that maps composite functions onto limited range SLMs is that this mapping is a nonlinear transformation. Thus, when composite functions are encoded, the individual functions of the composite function can effectively mix and produce noticeable sum and difference frequency harmonics. These undesired orders are quite evident for spot array generators from Bell Labs for photonic switching and from Teledyne Brown for structured light inspection systems. In optical correlators such mixing products can be incorrectly identified as false correlation peaks. This problem is illustrated in Fig. 5a for phase-only pseudorandom encoding. The diffraction pattern in the left column is the result of encoding the desired complex function by pseudorandom encoding and the right column is the result for encoding using the *phase-only filter* (i.e. each desired amplitude is encoded as unity, and each desired value of phase is exactly realized. J. L. Horner and P. D. Gianino, "Phase-only matched filtering," *Applied Optics*, 25, 3767-3772. (1984)). The harmonic nature of the mixing products is evident for the right column. For the left column, the speckle noise is uniformly spread over the diffraction plane and thus the intensity is substantially lower than the intensity of the individual noise harmonics found for the phase-only filter. Thus false peaks are less evident if pseudorandom encoding is used. Also, the non-uniformity of the 64 desired spots is 7 % rms for the pseudorandom encoded modulation but over 100 % for the phase-only filter for the simulated results in the first row of Fig. 5a. (This information on nonuniformity is shown in Fig. 5c. This graph will be described in Sec. 4.3.) Since mapping errors are transformed into a broadly spread and average low level noise pattern, pseudorandom encoded SLMs behave more like complex valued SLMs and suffer much less from undesirable nonlinear interactions. This is a decided advantage of random encoding. The level of performance achieved by pseudorandom encoding has been further improved through our development of a generalized algorithm.

4.3 The generalized pseudorandom encoding algorithm.[11,15] This algorithm improves the diffraction efficiency of composite functions and can run at near real time rates on conventional low-end digital processors. The algorithm adds two more steps to the original pseudorandom encoding algorithm:

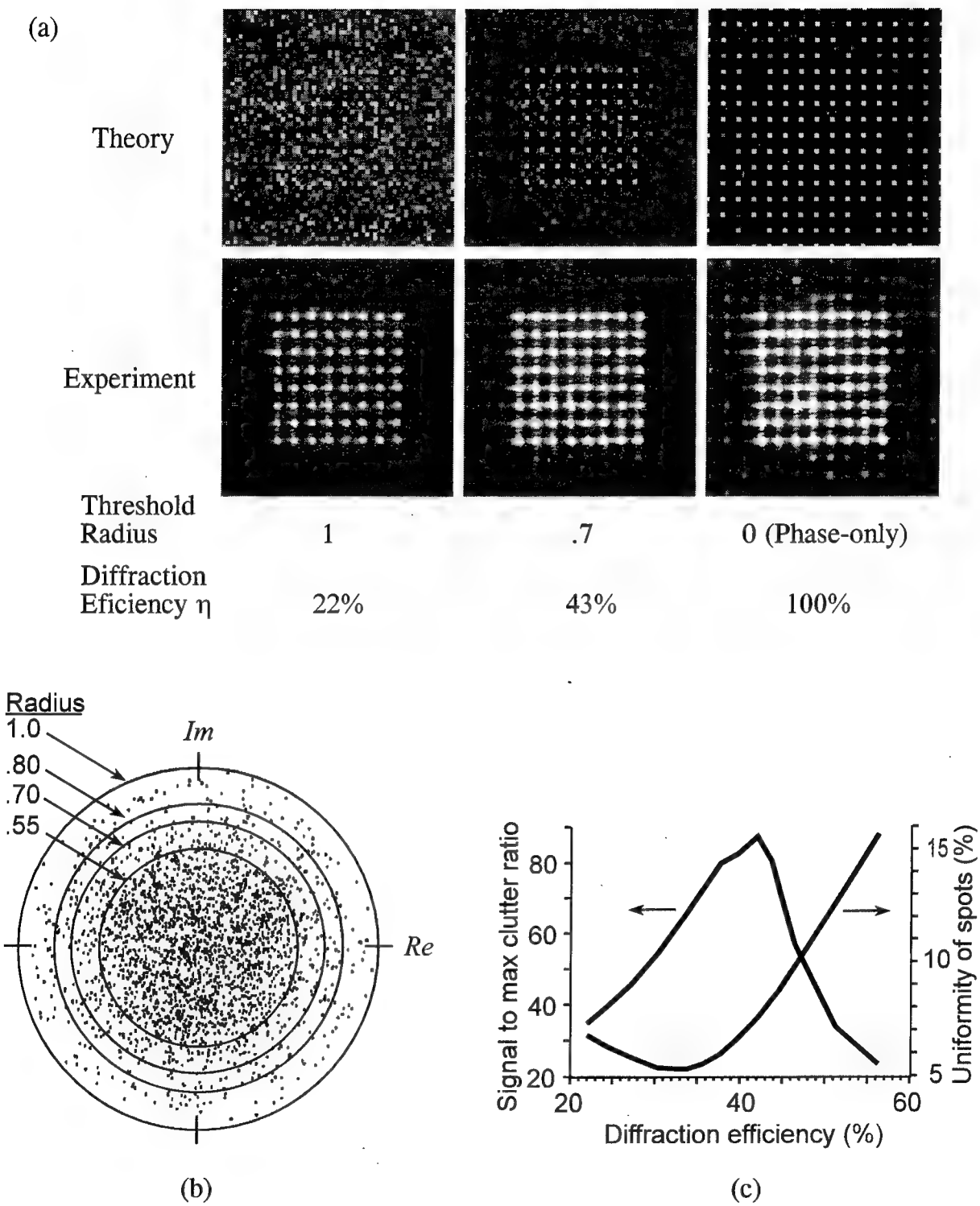


Fig. 5. (a) Diffraction patterns for various threshold radii; (b) scattergram of complex values that are encoded; (c) performance of spot arrays for decreasing threshold radius.

d) ***Specification of high efficiency encodings***.[14] The individual functions are added together with different phases. In particular we choose the values of phase using a random number generator. Even better performance is achieved by repeating this step a number of times and each time recalculating the diffraction efficiency η. The complex modulation having the highest diffraction efficiency is selected for pseudorandom encoding by steps a-c. (Sec. 3.1) The individual functions are specified in the SLM plane so that the addition of the functions, as well as the calculation of the diffraction efficiency are done without using the fast Fourier transform. This permits the calculation to be performed with a relatively small number of computations.

e) ***Partial encoding for balancing random and systematic errors***.[11,15] The complex values of the desired modulation that have amplitudes that are less than a specified value, referred to as the *threshold radius*, are pseudorandom encoded. (Fig. 5b illustrates the desired complex values, resulting from step d, and four possible threshold radii.) Complex values exceeding the threshold radius are encoded by specifying the corresponding phase-only filter (of Horner and Gianino, reference given above.) This phase-only encoding consists of setting each amplitude to the same level (that of the threshold radius) and keeping the phase angles identical to those of the desired complex values. Currently the optimal threshold radius is found iteratively. Steps a-c are repeated for different values of radius until a modulation is found that produces the best performance according to a specified criterion (e.g. diffraction efficiency, SNR, PNR, etc.) It appears that models, derived using eqs. (3),(4) and (10), can be developed to select a nearly optimal radius directly.

4.4 Results for the specific design of an 8 x 8 spot array generator.[13,15] The generalized design procedure was applied to design an 8x8 spot array generator. The results (both theoretical and experimentally demonstrated) are illustrated in Fig. 5. Fig. 5b shows the complex values comprising the desired filter. These values were found using the procedure described in step d. The diffraction efficiency η increased from under 5 % (all phases identical) to 22 % after performing this step. Next the threshold radius was adjusted over the range from 1 to 0 (step e). For this design we define the optimal encoding to be the one producing the maximum intensity diffraction spots to peak background intensity. This metric is designated as *signal to max clutter ratio* in Fig. 5c. It is maximum at a diffraction efficiency of 42 % which corresponds to a threshold radius of 0.7. (Unless otherwise stated these values are for the theoretical calculations.) Note that over this plotted range of η from .22 to .56 the radius decreases from 1 to .55 in a nearly linear fashion. The 64 spots were found to be most uniform in intensity for a threshold radius of .8 (5% standard deviation from uniform). The designs are nearly as uniform for radii of 1 and .8 (6.8 % and 7.5 % respectively.) The corresponding diffraction patterns (for various threshold radii) are show in the top row of Fig. 5a. These grayscale images of intensity are saturated (50x in first two columns and 33x in the third column) in order to bring out the background noise and to specifically illustrate that the middle image indeed has the lowest background noise. The experiment (using a Hughes LCLV) shows similar trends. While the left and center images do appear to have roughly the same level of background speckle, note that the diffraction spots more heavily saturate the CCD camera for partial encoding and thus the signal to background noise is higher in the experimental demonstration as well. It should be noted that

the experiment is only qualitatively correct. Further experimental characterization of this SLM and calibration of the optical experiment is required before a quantitative comparison is possible.

Explanation of improvement from step d. Since the phases of the diffracted spots are randomly selected, the complex modulation (the inverse Fourier transform of the desired diffraction pattern) looks quite similar to a speckle pattern. The speckle pattern has a much lower dynamic range in the values of intensity. Basically this reflects the fact that the peak intensity of a speckle pattern resulting from M randomly phased radiators is $1/M$ times smaller than for M coherently phased radiators.

Further increase in diffraction efficiency is possible by choosing one of several possible random phasings. The reason for this can be qualitatively explained with reference to Fig. 5b. This figure shows that there are fewer complex values with increasing radius. Pseudorandom encoding normalizes the complex values by the maximum complex value, as indicated by the radius of 1 in Fig. 5b. But the maximum value is in the tail of the statistical distribution of all amplitudes and thus its value can vary significantly from iteration to iteration of the design algorithm. There are, however, a large number of values of small amplitude, and due to this large number, the statistical distribution does not vary significantly from iteration to iteration. Therefore, the statistical fluctuations of the maximum intensity modulation value sets the diffraction efficiency. Thus the objective of step d is essentially a search for a set of complex values for which the maximum intensity is minimum.

Explanation of improvement for step e. The improvements resulting from partial encoding can be understood in terms of systematic and random types of errors. (A fuller discussion of systematic and random phase errors can be found in Sec. 5.1) Consider Fig. 5b. For a radius of unity the only errors are due to the randomness of the pseudorandom encoding algorithm. The encoding of amplitude as randomness of phase causes the lowest amplitude values to introduce the most noise. If the radius is reduced then the distances to the lower amplitude values are reduced, and this reduces the amount of random noise. However, a systematic error (in amplitude) is also introduced for each complex value outside the circle. The systematic error is the distance between a desired value and the threshold circle. Essentially any point outside the circle is reduced in amplitude (i.e. saturated) to the radius of the circle. The plot in Fig. 5c shows that there is a balance between systematic and random errors that produces the best performance. These results are even more suprising when one recognizes that for the best uniformity case ($\eta = .33$ or equivalently radius $= .8$) that only 5 % of the values are outside the circle and that for the best signal to clutter case ($\eta = .42$ or radius $= .7$) that only 12 % of the values are outside the circle. These results on partial encoding will be reported in Ref. 15.

4.5 Results for distortion invariant pattern recognition.[10,11] A simulation study was performed in which composite filters were designed to recognize an aircraft in any one of many distinct orientations. The filters (designed by Hassebrook) were encoded using steps a-c,and e. However, the particular composite filter design method did not permit the use of step d. The improvements, while improved over the phase-only filter, are not as dramatic as for the encoding of spot arrays.

The reasons include 1) the lower number of pixels in the filter SLM (64x64 as oppossed to 300x300); 2) that step d was not used to level the filter amplitudes; and 3) that the Fourier spectra of the individual training images (aircraft silhouettes), being low bandwidth, are much less level to begin with than the individual sinusoids used in designing spot arrays. Future investigations should focus on modeling to what degree these effects influence the quality of encoded composite filters. Models already developed by us (Sec. 5.0) already describe the effects of number of pixels and bandwidth on single function recognition filters. It appears likely that these models can be extended to reasonably model the performance of encoded composite filters.

4.6 Patterned diffuser arrays: Pseudorandom encoding applied to diffractive optics.[12,13] We have also begun developing an application of the pseudorandom encoding concept to the design and fabrication of fully-complex fixed pattern diffractive optics. The diffractive optic would consist of an array of pixels. Each pixel is a diffuser having a custom specified step height and a custom specified vertical roughness. The height corresponds to the desired phase and the roughness corresponds to a desired amplitude. Therefore, according to the physical model of scattering (Sec. 2.1 and Fig. 1) the amplitude of the specular component depends on the vertical roughness. Furthermore, the specular components from an array of diffuser pixels can be added together in the far-field, as illustrated in Fig. 2 to produce the desired diffraction pattern.

The advantage of the diffuser pixel over a pixel that represents a single random value of phase is that the higher spatial frequency of the diffuser produces a much broader pattern of diffuse scatter. This reduces the average noise power that can interfere with the desired (specular) diffraction pattern. In terms used by antenna designers, using a diffuser instead of a single pixel increases the *directivity* of the desired pattern with respect to the extent of the noise pattern. Say a diffractive optic is designed to have 100 μm square pixels and 2 μm roughness cells. The directivity gain is 50x50 or 2500! This provides a dramatic improvement in signal to noise over the designs presented above. Some preliminary simulations and comparisons that demonstrate the improvement are presented in Ref. 13. The diffuser array is a much better representation of the desired complex modulation than single phase per pixel devices. As with pseudorandom encoding in general, the design procedure is greatly simplified since the pixels are effectively complex valued.

These advantages of diffuser arrays may not at first appear compelling. For instance, many methods have been developed that produce an effective complex modulation by defining a superpixel as a cluster of individual pixels. However, it is the additional potential for reduced fabrication time and cost that makes the diffuser arrays quite appealing. The fabrication method we have been considering is to expose photoresist one diffuser pixel at a time. There is a range for which photoresist thickness (after development) is proportional to exposure energy intensity. This has already been used for laser writing of diffractive optics by raster scanning. For highly resolved patterns, such as the roughness functions we wish to pattern, the patterning can take several hours and the patterning equipment (typically an e-beam system) is quite costly. Instead we would pattern an entire pixel with a single exposure, step to the next pixel, and then expose with a new pattern. The exposure parameters consist of the average energy per unit area (which

corresponds to step height) and the spatial nonuniformity of the illumination (which corresponds to roughness).

There are many ways to achieve the desired pixel exposure pattern. Perhaps the simplest is to generate a rough pattern by passing laser light through a diffuser to produce laser speckle. The size of a roughness cell corresponds to the diameter of an individual speckle. A uniform intensity bias can be effectively created by time averaging speckle patterns together on the photoresist. Translating or spinning the diffuser in front of the laser beam with a specified velocity can be used to produce the desired offset. Additionally, a shutter or modulator is needed to control the total dose. In our lab we have found that there is negligible blurring of 2 μm speckle exposures, even when the recording was performed without a vibration isolation table. For this reason we consider the goal of 100 μm pixels with 2 μm roughness cells to be achievable in a laboratory setting. With standard commercial photoresists, lasers and translation stages it appears that it would be possible to construct a patterning system with less than $40,000 of equipment. This machine could easily pattern custom-specified complex-valued modulations having 16,000 pixels in 5 minutes. Dust would be non-critical and patterning could be performed without using a clean room.

4.7 Application to photomasks for grayscale photolithography.[13] If a diffuser is used as a mask in a projection printer, it will scatter some portion of the light outside the aperture of the imaging system. If the source illumination is incoherent then the resulting image will be a uniform intensity gray level illumination. A patterned diffuser array would then produce a grayscale image. This makes it possible to directly pattern three dimensional structures (such as micro-lenses and prisms) in photoresist in a single exposure. Thus the patterned diffuser array could serve as a master for mass-producing integrated and diffractive optics. References to related activities in grayscale photolithography at other institutions are also cited in Ref. 13.

4.8 Application to fully complex real-time programmable SLMs.[13] We have briefly considered if it is practical to produce a programmable SLM that consists of individually controllable diffusers. It may be possible to cascade a phase-only modulator (e.g. the FBDMD) with a random diffusing liquid crystal (e.g. polymer dispersed liquid crystal) to achieve fully complex modulation with only two electrical controls per pixel. This type of liquid crystal is already used in commercial applications, especially as electronically controllable privacy blinds for windows. The problems in developing such a device are limited by the current problems in developing good SLMs in general, and phase-only SLMs in particular. In this environment, the development of cascaded SLMs are viewed as being unnecessarily risky and costly. However, we believe that the addition of an electrically controllable diffuser would be one of the lowest risk approaches to consider for a tandem SLM.

4.9 Possibility of fully complex modulation for Fresnel diffraction. Because the roughness is so much higher in spatial frequency than the pixel sampling rate, the devices should also reasonably represent complex-valued modulation functions even into the Fresnel region (i.e. because the law of large numbers can be approached for the interference of the wavefronts from

even a few pixels. Refer to Fig. 2.)

5.0 NOISE PROPERTY USED TO ANALYZE OPTICAL PROCESSOR PERFORMANCE

Analyses of optical correlators have been developed[2-6] which are heavily based on the random phase analysis equations, eqs. (3),(4), (10). The particular emphasis of these analyses is to model the effect of phase errors in the filter plane SLM on the performance of the 4f correlator. Moreover, approximate forms that can be easily calculated. These equations, which give general insights into the performance of the correlator, are often to be preferred over numerically exact solutions that can only be found through detailed simulation. In these studies we have also shown that the performance levels predicted by the new models often agree quite closely with the results found through rigorous simulations, even when using using real-world imagery.

In our early work the filter plane SLM was assumed to be phase-only and only systematic phase errors were considered.[4] This was generalized to 1) include the case that the SLM can produce any value of amplitude between 1 and zero (e.g. Kumar and Hassebrook's fractional power filter) in addition to its phase modulation and 2) that the SLM produces both random and systematic phase errors.[6] The results of the generalized analysis are summarized following a review of the definitions of random and systematic phase errors.

5.1 Definitions of and sources of random and systematic phase errors in SLMs. Random phase errors $\delta\phi_r$ are simply phase errors that differ from the desired filter phase by a random amount. That is to say the phase errors are modeled as random variables. A simple example of a source of this phase error is thermal noise on the video signal used to program an electronically addressed SLM. Systematic phase errors $\delta\phi_s$ are nonrandom. They are described as a known function of the desired phase. A well known example of a source of systematic error is incorrect calibration between the video drive level and the desired phase shift. Another source of systematic error is simply that the SLM cannot produce a full 360° of phase shift. Thus (for a calibrated SLM) there is no phase error for desired phases less than the maximum phase shift and an increasing phase error for desired phase shifts in excess of the maximum phase shift. Another systematic phase error is quantization of either the address signal or the levels that the SLM can produce. In this viewpoint the binary phase-only filter is quantized to two levels and the phase error is due to the difference between the closest of the two binary phases and the desired analog phase. These examples show that many limitations of current SLMs can be viewed as systematic errors.

5.2 The resulting model.[6] The model developed evaluates the correlation peak formed in the output plane of the 4f optical correlator when the desired object is present on the input SLM. In this model the filter plane SLM is programmed to represent a distorted version of the complex spectrum of the input object. The distortion is due to the systematic and random phase errors inherent in the SLM and also the amplitude weighting programmed by this SLM. For an amplitude spectrum of the form $a'(f)$ the amplitude of the filter can be of the form $[a'(f)]^n$ for

any value of the exponent n. For n = -1,0,1 the SLM represents respectively the inverse, phase-only and matched filters. This general class of filters, for all values of n, is referred to as the fractional power filters (FPF). For this model the spectrum that is found at the output of the filter plane SLM is then $a\exp[j(\delta\phi_r+\delta\phi_s)]$ where $a=[a']^{n+1}$. It is easy to see that the FPF is a slight generalization of the phase-only filter (that was originally analyzed in Ref. 4). For this mathematical analysis each FPF designated by n, can be equivalently viewed as a phase-only filter being illuminated by an amplitude spectrum to the $n+1$ power. Once the wavefront at the output of the SLM is known, the resulting correlation peak that will form in the output plane is determined.

We have already shown how to evaluate random phase errors using eqs. (3),(4),(10). It is however, particularly interesting that the effect of systematic phase errors on correlation peaks is mathematically similar to that of random phase errors, to the extent that both systematic and random effects can both be calculated using averages. This is shown in Ref. 4 where the correlation peak amplitude $c(0)$ is written

$$c(0) = \int_{-B_f/2}^{B_f/2} a(f)\exp[j\,\delta\phi_s(f)]\,df \qquad \textbf{(14)}$$

where B_f is the spatial extent of the filter plane SLM. This equation is essentially of the form of B_f multiplied by the *temporal* or *time average*

$$\bar{g} = \frac{1}{T}\int_{-T/2}^{T/2} g(t)\,dt \qquad \textbf{(15)}$$

where f corresponds to the time coordinate and B_f corresponds to the period. Under many practical circumstances the temporal average can be equated to the ensemble average

$$\langle g\rangle = \int_{-\infty}^{\infty} g\,p(g)\,dg \qquad \textbf{(16)}$$

where $p(g)$ is the probability density function of g. This property [that eqs. (15) and (16) are equivalent] is referred to as *ergodicity*. In the analyses in Refs. 4 snd 6 this assumption was applied the spectrum at the output of the filter plane SLM $g\equiv a\exp(j\delta\phi_s)$. The assumption that the temporal average can be replaced with an ensemble average is also justified by the physics of diffraction from the filter plane to the correlator plane. The correlation plane is the Fourier transform plane of the filter plane. Thus the resulting correlation peak is a direct result of wavefront superposition [illustrated in Fig. 2 and mathematically expressed in eq. (14) for the amplitude on the optical axis] and the superposition is describable as an ensemble average.

To perform the analysis indicated by eq. (16) a joint probability density function $p(a,\delta\phi_s)$ must be specified. After looking at histograms of amplitude and phase spectra of a number of images of faces and tanks provided by J. L. Horner of Rome Laboratory, we felt justified in making the following assumptions: 1) that the amplitudes and the systematic phase errors could be modeled as random variables; 2) that the phase ϕ_s of the input spectrum (not the phase error) is uniformly

distributed over 360°; and 3) that each amplitude is statistically independent of each phase. Obviously the spectrum is non-random but, just as the histogram of an image is used to develop statistically-based models in image processing, we find it convenient and useful to model the image spectrum statistically. These two assumptions, while not precisely true for most images, appear to be adequately correct for modeling purposes and for developiong better insight into the performance of optical correlators. With these assumptions eq. (14) for the correlation peak amplitude is approximated as

$$c(0) \approx \frac{B_f \bar{a}}{2\pi} \int_{-\pi}^{\pi} \exp(j\,\delta\phi_s)\, d\phi_s \equiv B_f \bar{a}\, \overline{\exp(j\,\delta\phi_s)} \quad . \tag{17}$$

Eq. (17) shows that if there is no systematic phase error then the maximum peak amplitude of the correlation peak is obtained. The maximum is set solely by the spatial extent of the SLM and its average transmitted amplitude. Eq. (17) has been used to evaluate and compare several different types of systematic phase error.[3-6] We also compared these approximate results with exact simulations for specific images and found very close agreement.[3,4] We further found, by assuming that random and systematic phase errors are independent, that the peak correlation amplitude for combined systematic and random phase errors can be approximated as

$$c(0) \approx B_f \bar{a}\, \overline{\exp(j\,\delta\phi_s)}\, \left\langle \exp(j\,\delta\phi_r) \right\rangle \equiv B_f \bar{a}\, p_s\, p_r \tag{18}$$

where the symbols p_s and p_r are used as a shorthand to represent the amplitude loss due to systematic and random phase errors.[5,6] Eq. (18) is the main result needed to derive many different metrics of correlator performance. The one we focused on in the study was peak to noise ratio (PNR) (as defined by J. L. Horner in "Metrics for assessing pattern-recognition performance," *Appl. Opt.* **31,** 165-166 (1992)). This PNR metric is similar, but not mathematically identical to PNR as described in Sec. 3.2. Basically, PNR measures the peak intensity to the average intensity of the correlation plane background. The background level is calculated by omitting the peak intensity from the calculation of the average correlation plane intensity. Often signal to noise ratio (SNR) is defined similarly except that the correlation peak intensity is included in the calculation of the average correlation plane intensity. This difference is often ignorable except for the case where the correlation peak is very sharp and the noise level is low. We will show this below. In this report we also define PNR to be a ratio of intensities, rather than of amplitudes (as is done in Refs. 3-6). The amplitude ratio was originally chosen to reduce the dynamic range of plots of PNR versus amount of systematic phase error. The intensity ratio is used in this report to coincide with the PNR definition in Sec. 3.2 on pseudorandom encoding metrics.

The resulting expression developed in ref. 6 for PNR as a function of systematic and random errors is

$$\text{PNR} \approx \frac{(N-1)\, p_s^2\, p_r^2}{Z - p_s^2\, p_r^2} \tag{19}$$

where the approximation symbol is a result of dropping terms in the expression that are usually very small. The full expression is given in Ref. 6. The parameter

$$Z \equiv \overline{a^2} / \bar{a}^2 \tag{20}$$

is a normalized measure of the bandwidth of the amplitude spectrum at the output of the filter plane SLM. Z is maximum for a white signal. In this case the numerator (which is simply the variance of the amplitude spectrum) is equal to the denominator (which is the square of the average amplitude). Most image spectra have values of Z that have much smaller bandwidth. This leads to less sharp correlation peaks and lower PNR. Values of 6 or larger were used for the simulations in Refs. 3-6. For values of Z this large it is reasonable to ignore the term $p_s p_r$ in the denominator since it is always less than or equal to unity. Ignoring this denominator term shows the numerical similarity between PNR and SNR which is written

$$\text{SNR} = N p_s^2 p_r^2 / Z \quad . \tag{21}$$

Since for SLMs the number of pixels N is usually quite large, the difference between N and N-1 is of no practical significance. These expressions simply show the relationships between number of pixels, spectral bandwidth and phase errors in determining the quality of the correlation peak. While simple in form, the results often approximate the exact values of the metrics quite well. For instance, for the phase-only correlator that was digitally simulated in Ref. 4, eq. (19) for PNR and eq. (21) for SNR produced estimates that differed by no more than 10 % for all cases (consisting of 2 images x 6-7 levels of phase error x 2 different types of phase error). These mathematically simple results should be especially helpful to system engineers who are trying to quickly develop error budgets and performance analyses of correlator-based system, without performing detailed simulations. If it is decided that simulations should be performed, our performance analysis can serve as a guide to help in intelligently designing simulation studies.

5.3 Specific additional results and extensions. Specific phase errors (inluding systematic errors of quantization, binarization, linear phase error and saturated phase error; and random uniform and gaussian errors) were evaluated in closed form and their relative effects were compared in Ref. 6. Systematic quadratic phase errors were evaluated numerically in in Ref. 3. This specific result can be used to help decide whether it is necessary to compensate for phase shift that is proportional to the square of address voltage in FBDMDs. This problem was first described by J. L. Horner and P. D. Gianino in "Effects of quadratic phase error on correlator performance," *Appl. Opt.*, 31, 3876-3878, (1992). Our analysis indicates that the performance loss for not compensating the quadratic phase error is ignorable if the maximum phase error introduced is less than one quarter wavelength. In fact, this new *quarter wave criterion* for correlators appears to be a good rule of thumb for most types of systematic and random phase errors. However, this analysis is valid for there being only one type of phase error. Either eq. (19) or eq. (21) shows that a tighter bound is needed if there are multiple sources of phase error. Statistical error bounds on the PNR due to the random phase errors are also derived in Ref. 6 through the use of eqs. (3), (4), (10).

The methods applied in the development of the current model could be applied to the more involved modeling of amplitude-phase coupled SLMs, the mapping of composite, distortion invariant filters to the SLM, and the addition of noise to the input scene.

6.0 CONCLUDING REMARKS

This study grew out of the assumption that statistical models of scattering and diffraction from random rough surfaces could be used to gain insight into the performance of optical processors, and could be used to devise new approaches to performing signal processing operations with phase-only and amplitude-phase coupled SLMs. Both of these objectives were achieved.

The specific signal processing operation developed was the mapping of complex valued filters onto SLMs that do not produce all complex values through the procedure known as pseudorandom encoding. The method is especially useful for real-time processors because the encoding algorithm requires as few as one function calculation per SLM pixel to be encoded, and the calculation be done even faster using lookup tables. Several logical extensions to pseudorandom encoding were also begun and further developed during the study. These include the generalized pseudorandom encoding algorithm for improving the performance of composite filters, pseudorandom encoding of coupled amplitude phase SLMs, patterned diffuser array, low cost systems for fabricating patterned diffuser arrays and additional applications of patterned diffuser arrays to the mass production of three dimensional micro-optics by projection printing methods.

The performance analysis equations developed here provide a useful starting point for analyses of many optical processors in that they simplify the mathematics needed to evaluate diffraction from arrays of random phase and amplitude SLM pixels. These analyses were used to evaluate the performance of SLMs perturbed by piston and tilt errors, of correlators subject to random errors in the filter plane and of the performance limits of pseudorandom encoding. The insight developed by the statistical modeling also led us to the new result that the spectrum of a deterministic image can often be well modeled as a random signal for the purposes of evaluating the effects of systematic phase errors on correlators. This result even led us to the development of a model that combines the effects of random and systemic SLM phase errors and which provides information on the relative effects of either type of error. A more detailed summary of the findings is listed in Sec. 1.3.

There are many fruitful directions in which this work can be extended. Some are in progress, under funding from Federal agencies (Sec. 8.0). Many of these recommendations pertain to logical extensions of pseudorandom encoding and error analyses to the fields of optical processing, diffractive optics and photolithography. This research is all concerned with diffraction from one plane to another.

However, perhaps the most interesting, mathematically challenging and having the greatest potential payoff is the analysis and application of scattering from thick or distributed gratings. Many systems including volume holograms, three dimensional optical memories, distributed feedback lasers, acousto-optic tunable filters, submicron scale gratings and distributed reflector filters in fibers all work on the principle of scattering from multiple reflectors. There is a well known Fourier transform relationship between the strength of the scattering from the reflectors and the the wavelength selectivity (i.e. frequency response) of the grating. Full complex

weighting could be envisioned by including a spatially varying, random perturbation on top of a periodic grating structure. This could provide the ability to shape the frequency response of the structure. A pseudorandom encoded gratings could ultimately lead to significant reduction in the crosstalk between adjacent wavelengths in the acoustooptic tunable filter; extend the tuning range of distributed feedback laser diodes; and similarly improve any distributed grating structure.

However, it is well known that if there are too many scatterers along the propagation direction then multiple scattering can seriously affect the filtering properties of the grating, even if the individual scatterers are weak. For deterministic gratings, Kogelnik's coupled wave theory for thick gratings is widely used. The logical step then in applying pseudorandom gratings would be to develop the statistical equivalent of Kogelnik's coupled mode equations. The mathematics appears daunting, but we believe that closed form equations that are simple enough to provide physical insight should be possible to develop in many cases. A complete list of recommendations for future studies are listed in Sec. 1.3.

A general conclusion that we have drawn from this study is that while statistics is usually viewed as a difficult field, statistics actually simplifies modeling and reasonably describes the wavefront superposition inherent in optical (especially Fourier transform type) processors.

7.0 LIST OF PUBLICATIONS

In all, there have been 21 publications or presentations resulting from this study. At least 4 more, for which the studies are essentially complete, are anticipated to be submitted in the near future. Citations have been grouped first by subject area, and secondly in chronological order. Each citation of a full-length paper or patent is identified by the reference number used in the report. If available, a reprint of each has been attached. Reprints of future publications will be forwarded to the contract monitor as they become available.

7.1 Resulting from Contract Support to Date. These include 6 papers in refereed journals, 1 patent, 1 invention disclosure being prepared for patent application, 5 proceedings papers, and 8 talks. Note that Ref. 7 received a best research paper award from the American Society of Engineering Educators (ASEE).

Error Analyses

Random phase errors only

1. R. W. Cohn and R. J. Nonnenkamp, "Statistical Moments of the Transmittance of Phase-Only Spatial Light Modulators," in *Miniature and Micro-Optics: Fabrication and System Applications II*, U. Roychoudhuri and W.B. Veldkamp, eds., *Proc. SPIE 1751*, 289-297. (20 July 1992, San Diego, CA)

2. R. W. Cohn, "Random Phase Errors and Pseudorandom Phase Modulation of Deformable Mirror Spatial Light Modulators," in *Optical Information Processing Systems and Architectures IV*, Bahram Javidi, ed., *Proc. SPIE 1772*, 360-368. (21 July 1992, San Diego, CA)

R. W. Cohn and R. J. Nonnenkamp, "Speckle Generation by Phase-Only Spatial Light Modulators," *Symposia on Defects in SLMs and Their Remedy, Optical Society of America Annual Meeting*, Albuquerque, N.M., Technical Digest 23, TuS2, p. 64. (22 September 1992)

Systematic phase errors only

3. R. W. Cohn and J. L. Horner, "Limited Phase Modulation and its Effect on Phase-Only Correlation," in *Optical Pattern Recognition V*, D.P. Casasent, ed., *Proc. SPIE* 2237, 147-151. (5 April 1994, Orlando, FL)

4. R. W. Cohn and J. L. Horner, "Effects of Systematic Phase Errors on Phase-Only Correlation," *Applied Optics*, **33**(23), 5432-5439. (10 August 1994)

Combined random and systematic phase errors

5. R. W. Cohn and J. L. Horner, "Performance Models of Correlators with Random and Systematic Phase Errors," in *Advances in Optical Information Processing VI*, D.R. Pape, ed., *Proc. SPIE* 2240, 270-277. (7 April 1994, Orlando, FL)

6. R. W. Cohn, "Performance Models of Correlators with Random and Systematic Phase Errors," *Optical Engineering*, **34**(6), 1673-1679. (June 1995)

Pseudorandom Encoding

Fundamental theory, simulations and experimental demonstrations

R. W. Cohn and M. Liang, "Approximation of Full-Complex Filters by Pseudo-Random Phase-only Modulation," Invited Talk *Gordon Research Conference on Optical Signal Processing and Holography*, Plymouth, NH. (29 June 1993)

R. W. Cohn and M. Liang, "Approximating Fully Complex Spatial Modulation with Pseudo-Random Phase-only Modulation," *Symposium on Binary and Diffractive Optics. Optical Society of America Annual Meeting*, Toronto, Canada, Technical Digest 16, ThU2, p. 184. (7 October 1993)

7. R. W. Cohn and M. Liang, "Approximating Fully Complex Spatial Modulation with Pseudo-Random Phase-Only Modulation," *Applied Optics*, **33**(20), 4406-4415. (10 July 1994)

8. U.S. Patent 5363186, R. W. Cohn and M. Liang, "Method of Producing an Optical Wave with a Predetermined Optical Function." (7 November 1994)

R. W. Cohn, "Approximating Fully Complex Spatial Modulation with Pseudo-Random Phase-Only Modulation," 1995 Research Award Address at *ASEE Southeastern Region Annual Meeting* as part of ASEE Research Award for best 1994 refereed journal paper (Ref 5.) (Biloxi, Miss. 10 April 1995)

9. R. W. Cohn and M. Liang, "Pseudorandom Phase-Only Encoding of Real-Time Spatial Light Modulators," *Applied Optics*. (Accepted December 1995, Galley proofs attached)

Application to composite/SDF pattern recognition filters

10. L. G. Hassebrook, M. E. Lhamon, R. C. Daley, R. W. Cohn, and M. Liang, "Using Pseudorandom Phase-Only Encoding to Approximate Fully Complex Distortion-Invariant Filters," in *Optical Pattern Recognition V*, D.P. Casasent, ed., *Proc. SPIE* 2237, 204-211. (5 April 1994, Orlando, FL)

11. L. G. Hassebrook, M. E. Lhamon, R. C. Daley, R. W. Cohn, M. Liang, "Random Phase Encoding of Composite Fully-Complex Filters, *Optics Letters*, **21**(4), 272-274. (15 February 1996)

Extensions for fixed pattern diffractive optics

12. R. W. Cohn, A. A. Vasiliev, and K. M. Walsh, "Method and Apparatus for Patterning Microtopographic Relief Surfaces in Photoresist," U.S. Patent application (In preparation by Rome Laboratory, Hanscom AFB Patent Department. Submitted to Hanscom 27 February 1995)

R. W. Cohn, A. A. Vasiliev, D. L . Hill, W. Y. Liu and K. M. Walsh, "Patterned Diffusers in Photoresist for Full-Complex Modulation," *Symposium on Diffractive Optics Fabrication. Optical Society of America Annual Meeting*, Portland, OR, paper WB4. (13 September 1995)

13. R. W. Cohn, A. A. Vasiliev, W. Y. Liu, D. L. Hill, "Fully Complex Diffractive Optics Design via Patterned Microdiffuser Arrays," (Submitted to *JOSA A*, 7 February 1996)

Survey on Pseudorandom Encoding

R. W. Cohn, M. Liang, A. A. Vasiliev, W. Y. Liu, and K. M. Walsh, "Recent Advances in Pseudorandom Encoding of Full-Complex Modulation," Invited Talk *Gordon Research Conference on Optical Signal Processing and Holography*, Plymouth, NH. (28 June 1995)

ARPA Optoelectronics Program Reviews

R. W. Cohn, "Speckle Generation by Phase-only SLMs," *Advanced Research Projects Agency (ARPA) Optics Review*, Hilton Head, S.C. (10 February 1993)

R. W. Cohn, "Speckle Generation by Phase-only SLMs," Advanced Research Projects Agency Materials Technology Office (ARPA/MTO) Optoelectronics Program Review, Monteray, CA (17 June 1994)

7.2 Publications in progress that will acknowledge contract support. At least 3 refereed journal publications and 1 U. S. patent application are anticipated.

Pseudorandom Encoding

Spot array generation, theory and experimental via phase-only pseudorandom encoding

14. R. W. Cohn and M. Liang, "Implementation of Composite Functions on Phase-only Modulators by Methods of Pseudorandom Encoding," To be submitted to *Applied Optics*.

15. R. W. Cohn, W. Y. Liu, A. A. Vasiliev and S. B. Bidiwala, "Improved Design of Composite Filters by Partial Pseudorandom Encoding of Composite Functions onto Phase-only Modulators," To be submitted to *Applied Optics*.

Extensions for coupled amplitude-phase modulators

16. R. W. Cohn, "Method of Mapping Complex Valued Functions to Amplitude-Phase Coupled Spatial Light Modulators," Invention Disclosure to be submitted to Rome Laboratory.

17. R. W. Cohn and W. Y. Liu "Statistically Based Synthesis of Complex Functions for Amplitude Coupled Phase Modulators," To be submitted to *JOSA A*.

8.0 SUBSEQUENT FUNDING RELATED TO THIS STUDY

R. W. Cohn, "Testbed for Precision Characterization of Spatial Light Modulators and Optical Processors," DAAH04-93-G-0467, Army Research Office (ARO), $68,714 (1 September 1993 - 31 August 1994) *Equipment grant including 2Kx2K digital CCD camera and components necessary to construct a 2" aperture phase shift interferometer for characterizing SLMs.*

R. W. Cohn, D. L. Chenoweth, L. G. Hassebrook, K. M. Walsh, "Technology Base in Optical and Opto-electronic Correlators," NCCW-60, National Aeronautics and Space Administration (NASA) through Western Kentuky University, $375,000. (1 July 1994 - 14 July 1997)

Interdisciplinary study among four faculty investigators who are studying different aspects of optoelectronic correlators. Areas include scene and target modeling; effects of SLM properties on correlation performance; composite filter designs for phase-only and coupled amplitude phase SLMs; development of high level supervisory functions necessary for truly adaptive and automatic target recognition; development of lower cost fabrication processes for making custom filters and diffractive optics; and end-to-end models of correlation/recognition performance.

R. W. Cohn, "Influence of Non-Ideal Spatial Light Modulators on the Performance of Optical Processors," DAAH04-94-G-0358 Advanced Research Projects Agency (ARPA) through Army Research Office (ARO), $137,050. (20 September 1994 - 19 September 1997) *AASERT award supporting graduate research assistants on studies on SLM phase errors and applications.*

A. A. Farag, D. L. Chenoweth, R. W. Cohn, "Laboratory for Computer Vision and Image Processing," CDA-9422094, National Science Foundation (NSF), $75,000. (15 April 1995 - 31 March 1996) *Equipment grant of single-user Silicon Graphics workstations dedicated to large scale design, simulations and analyses of image processing and optical signal processing problems.*

9.0 LIST OF TECHNICAL CONTRIBUTORS TO THIS STUDY

Robert W. Cohn	Principal Investigator and Professor of Electrical Engineering. He was responsible for modeling, concept development, experimental design, project direction.
Minhua Liang	Post doctoral research scientist, fully supported by this contract. He was responsible for most of the simulations and experimental results from this study. He is currently an Optical Engineer with Neuristics Corporation, Baltimore, MD.
Shaad B. Bidiwala	M.S. candidate in Electrical Engineering who designed arrays of multiple spots using the pseudo-random phase-only approximation and first introduced concept of partial random encoding. Completed M.S. and is currently enrolled in UofL medical school.
Gregory Stout	M.S. candidate in Electrical Engineering. Assisted in installation and testing of frame grabbers and video projector. Completed M.S. and is currently a project engineer at PSI Energy.
Robert J. Nonnenkamp	M.S. candidate in Electrical Engineering. Implemented simulations on the effects of random phase errors on spot array generators and correlators. Completed M.S. and is currently an engineer at IBM.
Scott Burba	M.S. candidate. Performed electrical measurements of projector video driver circuits.
Laurence G. Hassebrook	Assistant Professor of Electrical Engineering, U. of Kentucky, Lexington, Ky. He was PI on the study described in Refs. 10,11. His portion of the work was originally started on his own initiative

	and unsponsored by any agency. Since then joint funding has been obtained from NASA for continuing these studies.
Michael E. Lhamon	Research assistant and Ph.D. candidate for Prof. Hassebrook.
Raymond C. Daley	Research assistant and Ph.D. candidate for Prof. Hassebrook.
Stanley A. Hertel	Electronic Engineer provided by UofL Graduate School. Developed assembly language utility routines for controlling a Data Translation frame grabber used in this study.

The research contributions of the following personnel are also noted in this report. While they were not supported on the study, they did contribute to extending research started under the study and which has been or will be reported in publications acknowledging sponsor support. These are:

Anatoly A. Vasiliev	Sr. Faculty Scientist. Dr. Vasiliev worked for the University from February 1995 through February of 1996. He currently is a Research Scientist at Physical Optics Corporation. Prior to this study he was a Scientist at Carnegie Mellon and before that he was a Senior Scientist at the Lebedev Institute, Moscow. He has published numerous journal articles, books and patents in the area of optical processing. He was responsible for the experimental demonstrations using the Hughes liquid crystal light valve in Fig. 5.
Wen Yao Liu	Professor, Department of Precision Intrument Engineering, Tianjin University. He was on Sabbatical at the University of Louisville for calendar year 1995. He performed computer simulations related to pseudorandom encoding of coupled amplitude phase SLMs.
David L. Hill	Undergraduate Research Assistant. Developing photoresist recording procedures.

10.0 REPRINTS OF PUBLICATIONS FROM THIS STUDY

Statistical moments of the transmittance of phase-only spatial light modulators

Robert W. Cohn and Robert J. Nonnenkamp

University of Louisville, Department of Electrical Engineering
Louisville, KY 40292

ABSTRACT

The monochromatic diffraction patterns from phase-only spatial light modulators, such as deformable mirror- and liquid crystal-SLMs, are routinely modeled using the Fast Fourier Transform. Random errors in phase have typically been evaluated by Monte Carlo analysis. This can require the averaging of numerous runs to find the expected value of intensity and its standard deviation. However, the pixel structure of current modulators, which allows each pixel to apply arbitrary piston or tilt modulation, greatly simplifies the form of the expectations. Specifically, we assume that the parameters of piston, and likewise tilt, describing the entire SLM transmittance are gaussian, independent and identically distributed random variables. Expressions for the expected value of the Fourier plane intensity spectrum and its standard deviation, and for propagation of the intensity pattern to any arbitrary observation plane using the angular spectrum of waves formulation are presented. Each expression reduces to a small number of Fourier transform operations that may be efficiently calculated by the FFT.

1. INTRODUCTION

Phase-only spatial light modulators (SLM) are desirable for application to reconfigurable interconnects, neural networks and correlators. They can be made from liquid crystal[1,2] and deformable micro-mechanical mirrors.[3,4] The mirror elements currently of most interest either displace parallel to the optical axis (piston) or angularly deflect (tilt). Thus they may be used to either retard or deflect an incident plane wave a given amount corresponding to the amount of electrostatic force applied by addressing circuitry . A piston element suspended by four flexure-beams was recently demonstrated that imparts up to 4π piston phase modulation at visible wavelengths.[5] Arrays of 128 x 128 flexure-beam mirrors integrated with a video addressing circuit are anticipated this year from Texas Instruments.

Because SLMs are analog devices, they are subject to errors and imprecision that can be the limiting factor in the performance of information processing systems. Errors cannot be perfectly controlled for a number of reasons. For example, the microfabrication process for deformable mirror devices (DMD) can leave unwanted material residues and residual strains that alters the hinge mechanical properties.[6] Addressing circuitry variations, noise in electronics, and charge recombination all contribute to uncertainties about the absolute deflection of a deformable mirror pixel.

We have previously observed effects of phase errors in inverted cloverleaf DMDs (four opposing cantilever beams per pixel).[7] In the photograph (Fig. 3 of that article) the brightest diffraction orders are surrounded by a hazy background illumination approximately four diffraction orders wide. The haze is undesirable in that it reduces the contrast and diffraction efficiency of the SLM. In Monte Carlo simulations in which phase errors were modeled as randomly perturbed mirror deflections, a large number of narrow, closely spaced intensity peaks arose which qualitatively matched the intensity footprint of the observed haze.

This diffuse background is often called speckle, which is observed when laser light scatters from rough random surfaces.[8] The current statistical models of speckle are most developed for rough and uncorrelated random surfaces. However, SLM surfaces are typically very smooth and structured. Rather than considering the SLM as a rough surface, it is reasonably modeled as a smooth surface perturbed by random parameters, e.g. piston, tilt, curvature, etc. The models need to be valid for any value of variance since precisely manufactured SLMs can have small phase errors, while binary phase encoding has and pseudo-random modulation is modeled as having large phase errors.

From this perspective we develop and present various statistical moments that describe the sensitivity of the diffraction patterns to random phase errors of arbitrarily programmed phase-only SLMs.

2. PHASE MODEL AND PIXEL TRANSMITTANCE

In this analysis we are primarily interested in pixels that are intended to perform either piston or tilt, but not both. It may however be the case, due to fabrication errors or the physical nature of an SLM that two or more motions may be present simultaneously (Fig. 1.) For instance, a piston-type pixel may have a random component of tilt; a liquid crystal light valve (LCLV) illuminated by a point source optical control could exhibit, in addition to piston, a parabolic phase modulation due to fringing electric fields across the thickness of the liquid crystal layer. Certain other phase errors may be ruled out. We will

0-8194-0924-3/93/$4.00

assume that cantilever-beam, torsion-beam and flexure-beam (pistoning) pixels are made from thick metal, and that they are deflected by straining the much thinner hinge metal. The pixels will then be treated as perfectly planar and unwarped but which may be tilted.

The equations and subsequent plots are for reflective SLMs and thus there is 4π of phase shift for one wavelength of physical displacement of the SLM surface. This leads to the model for the phase of the i'th pixel

$$\phi_i(x) = \frac{4\pi}{\lambda} \{ \psi_i' + \sin[\theta_i'(x-\alpha)] \} \approx \psi_i + \theta_i(x-\alpha) \tag{1}$$

This definition of phase is negative of the standard convention for describing the complex-valued description of waves (and is used to cancel the negative sign in the standard phasor of the complex-valued description.) Since available SLMs have small tilts (θ_i' less than 10 °), the small angle approximation to the sine has also been used. The parameter α in eq. 1 is included to model different hinge positions for cantilever and torsion type DMD pixels. For instance for a DMD mirror of width w setting α equal to 0 places the hinge at the pixel center which represents the torsion-beam pixel while setting α to w/2 represents a right connected cantilever beam.

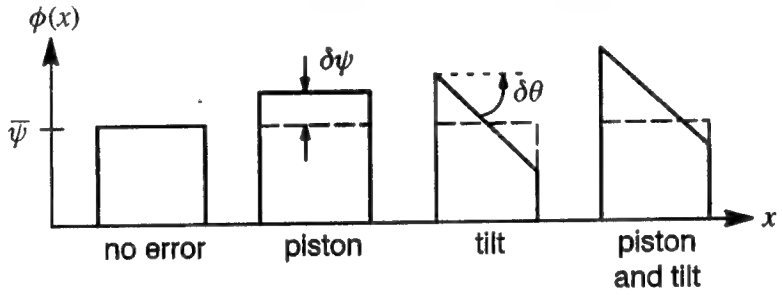


Fig. 1. Terms describing the phase of SLM pixels.

The expected value of a complex phasor with random arguments is easily found by using the definition of the characteristic function[9] of the random variable u

$$\langle \exp(j\omega u) \rangle = \int_{-\infty}^{\infty} \exp(j\omega u)\, p_U(u)\, du = 2\pi\, \mathcal{F}^{-1}\{p_U(u)\} \tag{2}$$

where $\mathcal{F}$ is the Fourier transform operator. If the parameters of piston ψ_i and tilt θ_i are independent and gaussian random variables then using eq. 2 the expected value of pixel transmittance $a_i(x)$ is found to be

$$\begin{aligned} \bar{a}_i(x) &= \langle a_i(x) \rangle = \langle r(x) \exp\{j[\psi_i + \theta_i(x-\alpha)]\} \rangle \\ &= r(x) \exp\left\{-\tfrac{1}{2}[\sigma_\psi^2 + \sigma_\theta^2 (x-\alpha)^2]\right\} \exp\{j[\bar{\psi}_i + \bar{\theta}_i(x-\alpha)]\} \\ &= g(x) \exp(j\bar{\phi}_i) \end{aligned} \tag{3}$$

where σ_ψ is the standard deviation of ψ_i, σ_θ is the standard deviation of θ_i, and where r(x) = rect(x/w) and g(x), the amplitude of the expectation of transmittance are defined to simplify subsequent expressions.

Eq. 3 shows that $\langle a_i(x) \rangle$ is reduced in amplitude over $a_i(x)$. For errors in piston only, the amplitude of transmittance is reduced uniformly across the aperture of the pixel, while for tilt errors there are position dependent changes in amplitude. Thus tilt apodizes the effective length (or area in two dimensions) of a pixel. For large values of σ_θ the pixel on average behaves as a point source centered on the hinge at $x=\alpha$. Thus random tilt reduces the effective area of the pixels and is anticipated to diffract light over a greater angular extent.

3. SLM TRANSMITTANCE AND ANGULAR SPECTRUM

The transmittance of an N-pixel, phase-only SLM is

$$t(x) = \sum_{i=1}^{N} a_i(x) = \sum_i r(x-x_i) \exp[j\phi_i(x-x_i)] \tag{4}$$

Any inactive area between the pixels has been treated as non-reflecting and the amplitude and phase of each pixel have been defined in local coordinates centered around x=0 and then shifted to pixel locations x_i. (This will prove useful in later equations (such as eq. 15) where autocorrelation operations translate all pixels back to local coordinates.) The expected value of SLM transmittance in eq. 4 under the assumption that all piston errors are identically distributed and all tilt errors are identically distributed is

$$< t(x) > = \sum_i < a_i(x) > = \sum_i g(x - x_i) \exp[j \bar{\phi}_i (x - x_i)] \tag{5}$$

This result would further require the assumption that the parameters are independent at this point if the pixels physically overlap in space (which would be possible if fringing from one pixel address voltage, affects a neighboring pixel.) In any case, the additional assumption of independent parameters is made in order to simplify the form of the higher order moments of transmittance. We are particularly interested in the expected intensity spectrum

$$< I(f_x) > = < T(f_x) T^*(f_x) > = \mathcal{F} \{ < t(x) \oplus t(x) > \} \tag{6}$$

and the expectation of the square intensity spectrum

$$< I^2 > = < T T^* T^* T > = \mathcal{F} \{ < t(x) \oplus t(x) \oplus t(x) \oplus t(x) > \} \tag{7}$$

where $T(f_x)$ is physically, the angular spectrum and mathematically, the Fourier transform of the transmittance t(x); and where we have used linearity of expectation and the Fourier transform to interchange these two operations in eqs. 6-7; and where $\oplus$ indicates the correlation integral

$$a(x) \oplus b(x) = \int a(x' + x)\, b^*(x')\, dx' \tag{8}$$

The standard result for standard deviation of intensity $\sigma_I(f_x)$ is then

$$\sigma_I^2 = < I^2 > - \bar{I}^2 \tag{9}$$

The simplifications resulting from statistical independence are illustrated by applying this assumption (in several successive steps) to eq. 6 as follows

$$\begin{aligned} \bar{I}(f_x) &= \sum_i \sum_j < A_i(f_x) A_j^*(f_x) > \\ &= \sum_{i \neq j} \sum < A_i > < A_j^* > + \sum_i < |A_i|^2 > \\ &= \sum_i \sum_j < A_i > < A_j^* > - \sum_i |< A_i >|^2 + \sum_i < |A_i|^2 > \\ &= |\bar{T}|^2 + \sum_i \left[< |A_i|^2 > - |< A_i >|^2 \right] \end{aligned} \tag{10}$$

This general result shows that the expectation nearly separates into $<T>^2$ except for where the transmittances of the same pixel overlap. A similar procedure, with a lot more algebra (for more details see related derivation in appendix C of ref. 9,) can be performed to find the second order moment of the intensity spectrum.

$$\begin{aligned} < I^2 > = & \sum_i \sum_j \sum_k \sum_l < A_i A_j^* A_k^* A_l > \\ & = 2 \left[\bar{I}^2 - |\bar{T}|^4 \right] + |\bar{T}^2 + \sum_i (< A_i^2 > - \bar{A}_i^2)|^2 \\ & + 4\, Re \left[\bar{T}^* \sum_i (< |A_i|^2 A_i > - < |A_i|^2 > \bar{A}_i^* + 2 |\bar{A}_i|^2 \bar{A}_i - 2 < |A_i|^2 > \bar{A}_i) \right] \\ & + \sum_i \left[< |A_i|^4 > - 6 |\bar{A}_i|^4 + 8 < |A_i|^2 > |\bar{A}_i|^2 - |< A_i^2 >|^2 - 2 < |A_i|^2 >^2 \right. \\ & \qquad \left. + 4\, Re (< A_i^2 > \bar{A}_i^{*2} - < |A_i|^2 A_i > \bar{A}_i^*) \right] \end{aligned} \tag{11}$$

Using the assumptions of identical gaussian statistics for the parameters of tilt and piston, along with the linear phase model of eq. 1 will lead to significant simplification of eqs. 10-11. With the additional simplification that $\alpha = 0$ in the phase model these equations become

$$\bar{I}(f_x) = \mathcal{F}\left\{ \bar{t}(x) \oplus \bar{t}(x) + [g_2(x) - p\, g_{11}(x)] \sum_i e^{j\bar{\theta}_i x} \right\} \tag{12}$$

$$< I^2(f_x) > = 2\left[\bar{I}^2 - |\bar{T}|^4 \right] + \left| \bar{T}^2 + T_A \right|^2 + 4\,Re\left[\bar{T}\, T_B^* \right] + G_A$$

where

$$\begin{aligned} t_A(x) &= \sum_i \left[p^2 g_2(x - 2x_i) - p\, g_{11}(x - 2x_i) \right] \exp\left\{ j[\, 2\bar{\psi}_i + \bar{\theta}_i(x - 2x_i)\,] \right\} \\ t_B(x) &= p^{\frac{1}{2}} \sum_i \left[g_3(x - x_i) - (2 + p^2) g_{21}(x - x_i) + 2p\, g_{11}(x - x_i) \right] \exp\left\{ j[\, \bar{\psi}_i + \bar{\theta}_i(x - x_i)\,] \right\} \\ g_A(x) &= \left[g_4(x) + 4p\, g_{31}(x) - (2 + p^4) g_{22}(x) + 4p(2 + p^2) g_{211}(x) - 6p^2 g_{1111}(x) \right] \sum_i \exp(j\bar{\theta}_i x) \end{aligned} \tag{13}$$

and where the g-functions and constant p are defined as

$$\begin{aligned} p &= \exp(-\sigma_\psi^2) \\ g_{ijkl}(x) &= g_i(x) \oplus g_j(x) \oplus g_k(x) \oplus g_l(x) \\ g_n(x) &= \left[r(x) \oplus \cdots \oplus r(x) \right]_n \exp(-\tfrac{1}{2}\sigma_\theta^2 x^2) \end{aligned} \tag{14}$$

and n indicates an n-term correlation (n-1 integrals.) In the case of piston-only pixel modulation the equations further reduce to

$$\bar{I}(f_x) = |\bar{T}(f_x)|^2 + N q\, R^2(f_x) \tag{15}$$

$$< I^2(f_x) > = 2\left[\bar{I}^2 - \tfrac{N-2}{N} |\bar{T}|^4 \right] - \tfrac{4}{N}\left[|\bar{T}|^2 + \tfrac{N}{2} q^2 R^2 \right]^2 + \left| \bar{T}^2 - p\, q\, T_A \right|^2$$

where

$$t_A(x) = \sum_i g_{11}(x - 2x_i) \exp(j2\bar{\psi}_i) = r(x) \oplus r(x) * \sum_i \delta(x - 2x_i) \exp(j2\bar{\psi}_i) \tag{16}$$

where "*" is the convolution operator and where

$$\begin{aligned} q &= 1 - p \\ R(f_x) &= \mathcal{F}\{ r(x) \} \end{aligned} \tag{17}$$

Remembering the definition that r(x)=rect(x/w) then $R(f_x)$ is a sinc function.

3.1 Comments on the form of the equations

Referring to any of the expected intensity spectra (eqs. 12,15 or even 10) the equations are seen to reduce to two terms. The first term expresses the coherent or "specular" reflection from the surface and contains the coherent signal intended from the SLM. This contribution to the intensity spectrum is $|\mathcal{F}\{<t(x)>\}|^2$ and is easily calculated using the FFT. The expected transmittance of each pixel in the SLM is reduced as the standard deviation in piston and tilt is increased according to eq. 3. The second term in the SLM plane is only non-zero over a width of 2w (w being the width of a single pixel). The Fourier transform of this narrow impulse adds a broad pedestal to the coherent intensity spectrum. Terms comparable to this for theories of rough surfaces are referred to as "diffuse halo".[10] They are considered to be essentially incoherent, containing none of the information t(x) that is modulated onto the SLM. The incoherent spectrum is bandlimited by the transform of the rectangular r(x) functions that describe the finite width of the pixel. Note especially in eq. 15 that the pedestal is most pronounced when the standard deviation of piston is largest, i.e. when q equals unity. At this point p also equals zero so the coherent spectrum vanishes. Experimentally observing halo, we would actually expect to see a grainy spatial variation which is referred to as speckle. If the individual pixels were not independent (but selected from a correlated stationary process) their correlation would tend to broaden the impulse term in the transmittance, and consequently narrow the incoherent portion of the angular spectrum.

The following three subsections compare these equations with Monte Carlo simulations by considering: 1) computational efficiencies, 2) calculated results, 3) histograms of repeated Monte Carlo runs.

3.2 Implications for Computer Modeling

Eqs. 12-13 for tilt and piston, or 15-16 for piston alone are written in a form to suggest efficient computer calculations of the angular spectrum (or Fraunhofer diffraction pattern.) The $<t(x)>$ and the g-functions may be directly calculated in the space domain and then Fourier transformed by the FFT. The impulse terms, $g_A(x)$ and the pedestal term in eq. 12 have so few data points that the discrete Fourier transform (DFT) can often be computed with far fewer operations than the FFT. At most, eqs. 12-13 require five FFT's. If FFT's are used to perform the nine distinct g-functions, they need not require many calculations because of their short length. Eq. 13 (specifically t_A) requires non-zero samples over twice the extent of the SLM. Thus, Monte Carlo methods can take advantage of using half as wide an FFT. In our examples, however, in order to have adequate resolution in plotting narrow diffraction orders, the same size FFT was used for both.

We have coded these eqs. 12-13 and compared them with Monte Carlo simulations in terms of accuracy and computational efficiency. Fig. 2. shows these results plotted against m the number of Monte Carlo diffraction analyses run incorporated in the estimates of expected value and standard deviation of efficiency. For this case there is .1 λ standard deviation of piston error and the statistical averages are calculated at the peak of the zero-order diffraction pattern resulting from a nominally unmodulated 128 pixel SLM. While 20 Monte Carlo runs takes about the same amount of time as calculating eqs. 12-13, the accuracy of the standard deviation is not very good. Specifically, Fig. 2a. shows the relative error between the Monte Carlo estimate and the statistical moment. These errors are calculated from the mean and variance of 30 m-run Monte Carlo estimates of σ_I each of m runs. The upper curve is the standard deviation of the Monte Carlo estimate of σ_I normalized by σ_I. Over 200 runs are needed to reduce the standard deviation of the Monte Carlo estimate of σ_I to under 10 % relative error. The lower curve shows the relative error in bias between the estimate of σ_I and σ_I relative to σ_I. Each curve is apparently converging to our equations for the statistical moments. The expected value of intensity (not shown) was found to be within .1 % relative error for averaging 100 Monte Carlo runs. Away from the main peak the relative errors increase and more averaging is needed to obtain equivalent accuracy. The convergence of the Monte Carlo simulations to eqs. 12-13 (as indicated in Fig. 2) convinces us of their correctness for piston errors. This accuracy analysis has not yet been performed for the case of tilt errors; however, we have, to this point observed that single Monte Carlo runs qualitatively track the expected intensities and their sigma limits.

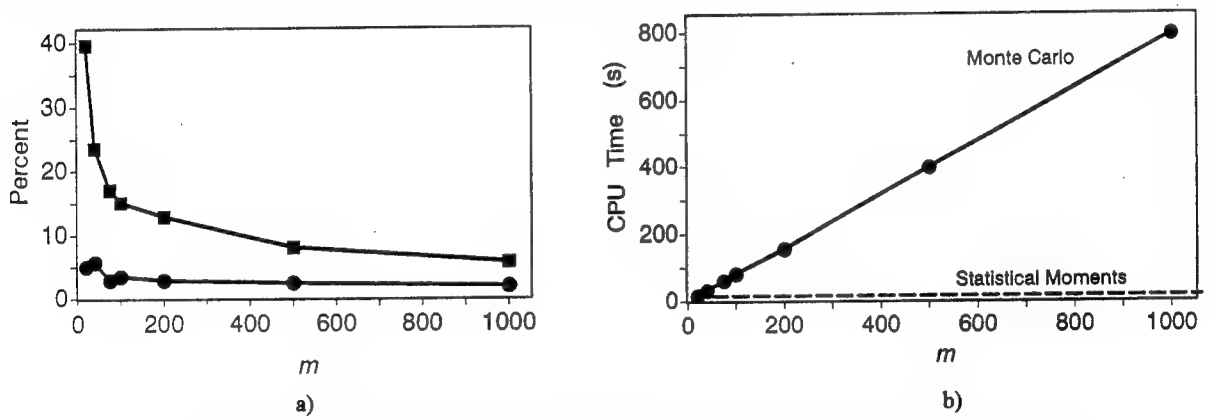


Fig. 2. Comparison between Monte Carlo analysis and statistical moment equations in terms of a) accuracy and b) computational efficiency.

3.3 Specific Examples

Figs. 3-5 plot single Monte Carlo runs on top of the corresponding expectation of intensity, and the $\pm\sigma_I(f_x)$ limits. In all cases these results are for a 128 pixel row with 100% duty cycle (fill factor) pixels, and there are nine FFT samples per pixel. Fig. 3 is the result if all pixels are identically deflected except for random piston error of standard deviation .1 λ. It may also be viewed as the correlation peak resulting due to residual phase-matching errors in the filter plane. The diffraction peak is reduced by a factor of .2 from the error-free case, but it is clearly distinguished from the sidelobes. Note that the negative one $\sigma_I(f_x)$ limit is zero (not negative) over most of the sidelobe region. Fig. 4 is the result for a diffraction grating that was designed by simulated annealing procedure to generate 4 equal-intensity, equally spaced spots. The design procedure was similar to that of Dames with the exception that it used eight phase levels rather than two.[11] The diffraction pattern in Fig. 4 corresponds to illuminating 4 periods of the 32 pixel design. The diffraction pattern may also be viewed as the response of a correlator to four identical, equally-spaced targets. The standard deviation of piston error is again .1 λ. With the light shared

between four diffraction peaks, the reduced energy into the desired diffraction peaks leads to severe interference from the sidelobes. Fig. 5 is again the case of all pixels identically deflected but with piston error of .3 λ. The standard deviation of intensity is essentially equal to the expected value of intensity, so that the negative standard deviation limit is the horizontal axis. The intensity of the erratic peaks tracks the broad sinc^2 pedestal. This is the envelope predicted by the pedestal term in eq. 15. Fig. 6 is the result for the case of 1 λ / w standard deviation of tilt (i.e. 1 λ total declination across the extent of the pixel.) In these figures the first diffraction order would be seen at 64 units of spot resolution for pixel duty cycles less than 100%

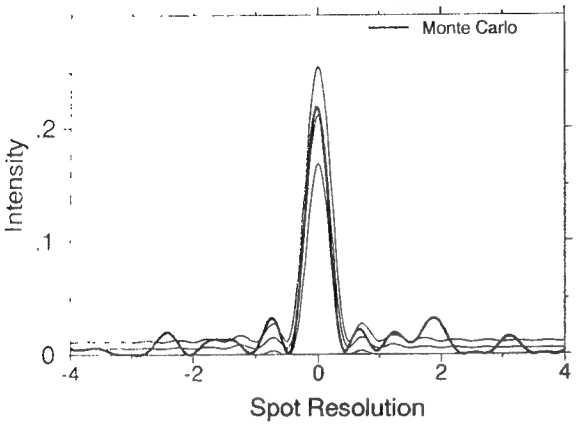


Fig. 3. Analysis of a nominally unmodulated phase grating with piston error of standard deviation .1 λ.

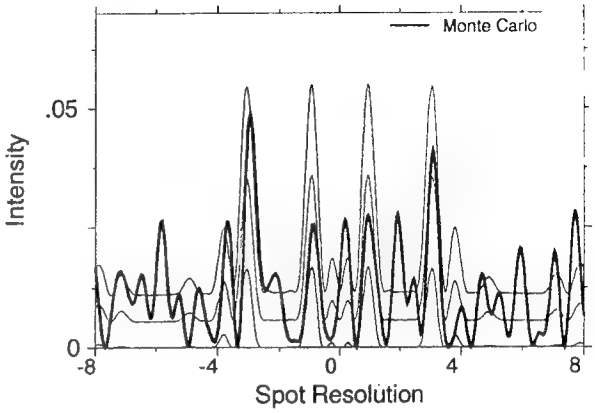


Fig. 4. Analysis of a four spot generating kinoform with piston error of standard deviation .1 λ.

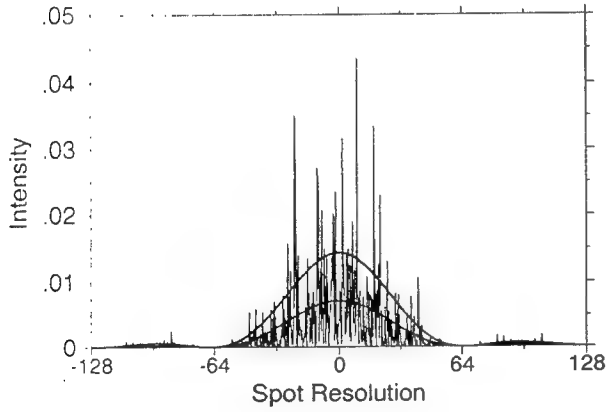


Fig. 5. Analysis of nominally unmodulated phase grating with piston error of standard deviation .3 λ.

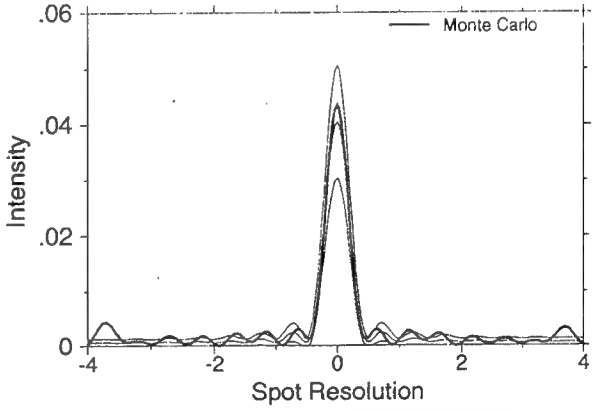


Fig. 6. Analysis of nominally uniform phase grating with tilt error of standard deviation 1 λ / w.

3.4 Histograms

From the large number of Monte Carlo runs used to generate Fig. 2 it was also possible to form histograms of the probability density function (PDF) of the intensity of the diffraction pattern. Results for the peak ($f_x=0$) are shown in Fig. 7a,b,c and the RMS errors of several least squares fits are shown in Fig. 7d. For small piston errors the gaussian gives the best fit and for large piston error the exponential gives the best fit. Neither fits the histogram for intermediate values of piston error. Similar results were found at the sidelobes, with the only difference being that much lower levels of piston error were required for the histogram to take on the exponential shape.

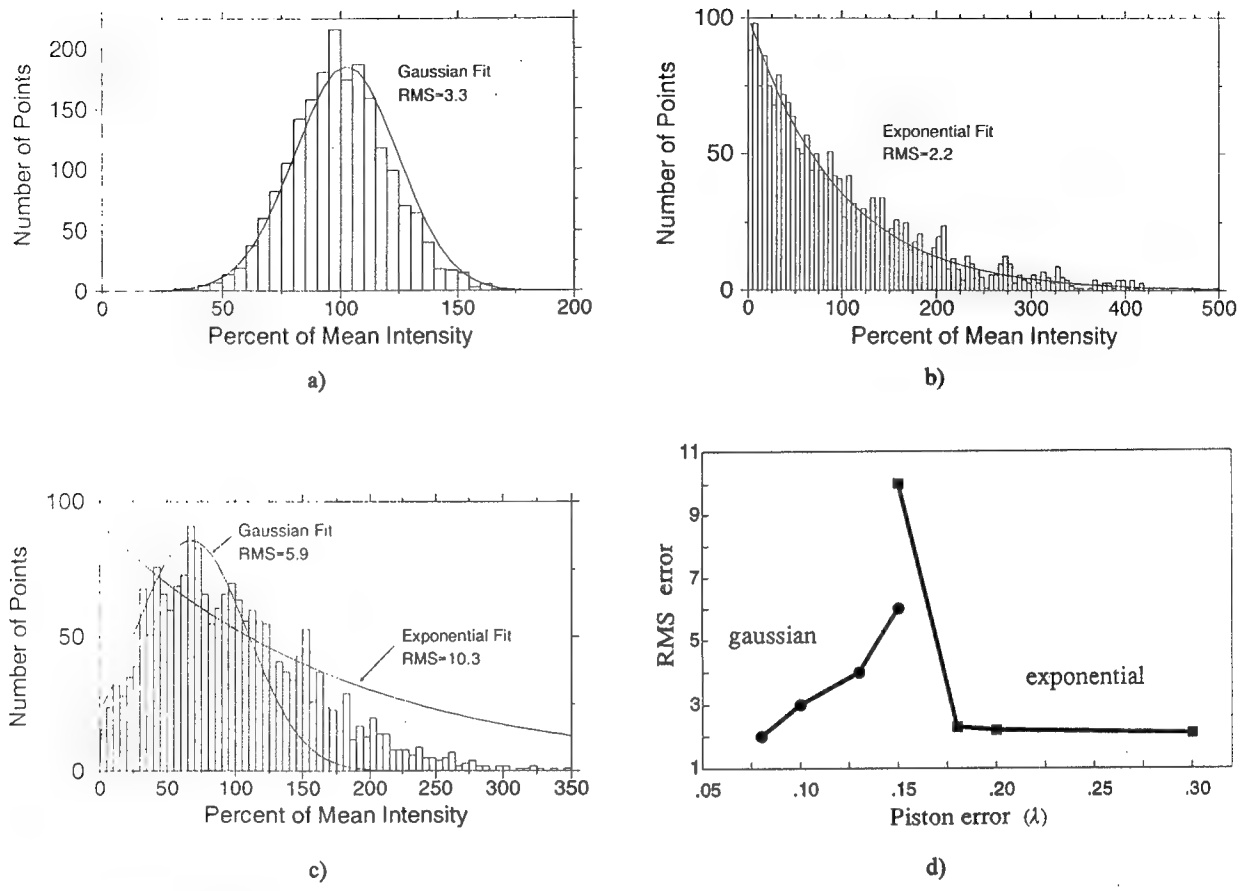


Fig 7. Histograms and their least squares fits for piston error of standard deviation a) .1 λ b) .2 λ c) .15 λ and d) fitting error *vs.* σ_I.

These histograms suggest the modified Rician functions.[8] The Rician family arises in describing the amplitude of complex random phasor sum added to a constant amplitude, and the family is parameterized as a function of incoherent to coherent magnitudes. The modified Ricians are simply the PDFs that result from the transformation of the Rician PDFs from random variables of amplitude into intensity. With no coherent background, the modified Rician is the exponential PDF. The standard deviation of an exponential PDF is equal to its expected value, which agrees with our earlier observations that the lower σ_I bound is often zero in the sidelobe regions or when the piston error is large. When the piston error is small the modified Ricians PDF can be shown to be well approximated by a gaussian PDF with a mean equal to the intensity of the coherent background, and a variance that is equal to twice the product of the background intensity multiplied by the expected intensity of the random phasor sum. At least for the case of piston errors only, the similarity of the modified Ricians to our histograms suggests that it may be possible to closely estimate the actual PDF based on knowledge of the expected value of intensity, the portion of the intensity due to the pedestal term, and the standard deviation of intensity. Knowledge of the PDF might then be employed in the determination of the probability of correctly identifying the presence of correlation peaks.

These equations for intensity and standard deviation of the intensity diffraction pattern should prove useful in determining the sensitivities of monochromatic diffraction patterns to random errors in phase-only SLMs with simultaneous reductions in simulation times.

4. GENERAL DIFFRACTION ANALYSES

There are several valuable physical interpretations of the Fourier transform in scalar diffraction theory: 1) as the primary component of Fraunhofer diffraction, 2) as the Fresnel diffraction from the object plane to the focal plane of a thin lens, 3) as the Fresnel diffraction from the Fourier plane to the correlation plane of an optical correlator, and 4) most generally as the angular spectrum of plane waves (ASW) of the object plane transmittance.[12] The last interpretation allows straightforward diffraction calculations to any plane by multiplying the object plane spectrum by a diffraction kernel and inverse transforming:

$$t(x,y,z) = \mathcal{F}^{-1}\left\{ T(f_x,f_y)\exp\left(j\,2\pi\frac{z}{\lambda}\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2}\right)\right\} = \Re\left\{ t(x,y,0)\right\} \tag{18}$$

where $\Re$ has been used to indicate the ASW operator and which is mathematically equivalent to the Fresnel-Kirchhoff diffraction integral.[13] The field t(x,y,0) is also written as t(x,y), or as t(x) in the one-dimensional equations given in this paper. The ASW formula when implemented using the FFT is often referred to by the name beam propagation method (BPM).

This section considers the generalization of the statistical moments of phase-only SLM diffraction patterns from the Fourier plane only to a plane at any arbitrary distance z from the SLM. The expected intensity at any plane follows by taking the expectation of the square of eq. 18 and rearranging using the correlation property of the Fourier transform as follows

$$\begin{aligned} <|t(x,z)|^2> &= \left\langle \left| \mathcal{F}^{-1}\left\{ T(f_x)\exp[\,j2\pi\frac{z}{\lambda}\sqrt{1-(\lambda f_x)^2}\,]\right\}\right|^2\right\rangle \\ &= \mathcal{F}^{-1}\left\{ \left\langle T(f_x)\exp[\,j2\pi\frac{z}{\lambda}\sqrt{1-(\lambda f_x)^2}\,]\oplus T(f_x)\exp[\,j2\pi\frac{z}{\lambda}\sqrt{1-(\lambda f_x)^2}\,]\right\rangle\right\} \\ &= \mathcal{F}^{-1}\left\{ \int \Gamma_T(f_x';f_x)\exp\left(j2\pi\frac{z}{\lambda}\left[\sqrt{1-\lambda^2(f_x'+f_x)^2}-\sqrt{1-(\lambda f_x')^2}\right]\right) df_x'\right\} \end{aligned} \tag{19}$$

where

$$\Gamma_T(f_x';f_x) = < T(f_x'+f_x)\,T^*(f_x') > = \mathcal{F}\{ < t(x)\exp(-j2\pi f_x x)\oplus t(x) > \} \tag{20}$$

is a statistical self correlation functionof the angular spectrum $T(f_x)$. Γ_T could also be called the spatial (as opposed to temporal) self coherence or the mutual intensity J_T.[9] The second argument of the correlation indicates that in general the integral in eq. 19 space variant and not practically simplified by applying Fourier transform properties. However, for the case of piston and tilt errors as described by eq. 1 the coherence function becomes

$$\begin{aligned} \Gamma_T(f_x';f_x) = \bar{T}(f_x'+f_x)\bar{T}^*(f_x') + \mathcal{F}\Big\{ \Big(&[\,r(x)\exp(-j2\pi f_x x)\oplus r(x)\,]\exp(-\tfrac{1}{2}\sigma_\theta^2 x^2) \\ &- g(x)\exp(-j2\pi f_x x)\oplus g(x)\Big)\sum_i \exp[\,j(\bar{\theta}_i x - 2\pi f_x x_i)\,]\Big\} \end{aligned} \tag{21}$$

4.1 Implications for computer modeling

The first term and the third (the correlation containing two g(x) functions) considerably simplify eq. 19. Each results in a correlation integral inside the Fourier transform operation. Thus the magnitude squared of the ASW operation on each term (see eq. 18) generates the corresponding two terms in eq. 19. The second term, the correlation containing two r(x) functions multiplied by a gaussian, does not simplify. While the calculation of the correlation function Γ_T is not a large part of the total numerical calculation, because the function does not separate in the Fourier plane, a two-dimensional ASW integral must be performed in order to solve eq. 19. Furthermore, if each tilt $<\theta_i>$ is a distinct value, then the ASW spectrum is translated and a different ASW analysis is performed for each pixel. Thus this equation simplifies somewhat if $<\theta_i>$ is unbiased or at least constant for all i. Specific approximations are also needed to simplify the second term.

For the case of piston errors only, eq. 21 reduces to

$$<|t(x,z)|^2> = |\Re\{\bar{t}(x)\}|^2 + q\,|\Re\{r(x)\}|^2 * \sum_i \delta(x-x_i) \tag{22}$$

This is especially easy to calculate. Eq. 22 suggests one approach for arbitrary tilt models by subdividing each pixel into multiple segments each with successively increasing amounts of piston. In this case however, the subelements are statistically correlated and this must be further accounted for.

5. CONCLUSIONS

We have presented expressions that describe the effects of random phase errors on the diffraction patterns of phase-only spatial light modulators. These equations extend the standard FFT-based deterministic models of diffraction and can be performed with 2 to 5 times the number of FFT operations as the deterministic case, which can be substantially faster than Monte Carlo estimates of the mean and variance of diffraction patterns. We also observed that histograms of piston only modulators track the modified Ricians probability density functions that are associated with laser speckle added to a coherent background, and that it may be possible to determine the correct pdf at every point of the diffraction pattern from our equations for mean and variance alone. While only one dimensional equations have been presented, they may be easily extended to two dimensional cases if it is physically reasonable to assume that tilts in x and y are independent of each other.

These equations may be applied to any phase-only modulators including optical interconnects and matched filters. In the design of binary optical fanouts there can be many local solutions.[14] It may be possible to incorporate the knowledge of the sensitivity of each solution to random phase errors into the design procedure so that fewer local solutions need to be examined. Some examples and extensions of these equations with specific application to optical correlators is presented in a companion paper.[15]

6. ACKNOWLEDGEMENTS

This work was supported in part by DARPA, through RADC, on contract F19628-92-K-0021.

7. REFERENCES

1. J. Amako and T. Sonehara, "Kinoform using an electrically controlled birefringent liquid-crystal spatial light modulator," *Appl. Opt.*, 30, 32, 4622-28 (10 November 1991). and references therein.
2. K. Lu and B.E.A. Saleh, "Complex amplitude reflectance of the liquid crystal light valve," *Appl. Opt.*, 30, 1374-78 (1991).
3. L.J. Hornbeck, "Deformable-mirror spatial light modulators," *Proc. SPIE* 1150, 86-102 (1989).
4. R.M. Boysel, "A 128 x 128 frame-addressed deformable mirror spatial light modulator," *Opt. Eng.*, 30, 9, 1422-27 (Sept 1991).
5. R.M. Boysel, J.M. Florence, and W.R. Wu, "Deformable mirror light modulators for image processing," *Proc. SPIE* 1151, 183-194 (1989).
6. W.R. Wu, R.O. Gale, L.J. Hornbeck, and J.B. Sampsell, "Electro-optical performance of an improved deformable mirror device," *Proc. SPIE* 825, 24-31 (1987).
7. D.A. Gregory, R.D. Juday, R. Gale, J.B. Sampsell, R.W. Cohn and S.E. Monroe, Jr., "Optical characteristics of a deformable mirror spatial light modulator," *Opt. Lett.*, 13, 1, 10-12 (January 1988).
8. *Laser Speckle and Related Phenomena*, second edition, J.C. Dainty, ed., Springer, Berlin (1984).
9. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York (1965).
10. J.W. Goodman, *Statistical Optics*, Wiley, New York (1985).
11. M.P. Dames, R.J. Dowling, P. McKee, and D. Wood, "Efficient optical elements to generate intensity weighted spot arrays: design and fabrication," *Appl. Opt.*, 30, 19, 2685-91 (1991).
12. J.W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, San Francisco (1968).
13. R.J. Collier, C.B. Burckhardt, L.H. Lin, *Optical Holography*, pp. 583-586. Academic Press, New York (1971).
14. F.B. McCormick, *"Generation of large spot arrays from a single laser beam by multiple imaging with binary phase gratings," Opt. Eng.*, 28,4, 299-304 (April 1989).
15. R.W. Cohn, "Random phase errors and pseudo-random phase modulation of deformable mirror spatial light modulators," *Proc. SPIE*, V. 1772-34 (1992).

Random phase errors and pseudo-random phase modulation of deformable mirror spatial light modulators

Robert W. Cohn
University of Louisville, Department of Electrical Engineering
Louisville, Kentucky 40292

ABSTRACT

Piston-only spatial light modulators, such as the flexure-beam deformable mirror device, hold great promise for real-time optical processors because of their ability to accurately match the phase of arbitrary laser images. Non-ideal devices may have uncontrolled phase errors which can limit performance. A statistical analysis has been developed which models the effect of random piston and tilt errors on the diffraction pattern of phase-only SLMs. A slight modification of these equations describes the performance of these SLMs in the phase-only correlator and the phase-only correlator modulated by pseudo-random sequences. Results on correlator diffraction efficiency versus amount of phase error are presented. For the specific case of a binary phase-only matched filter a diffraction efficiency of 40.5 % is found.

1. INTRODUCTION

Among phase-only spatial light modulators, the flexure-beam deformable mirror device (FBDMD) under development by Texas Instruments[1] appears the most promising for optical information processing applications, especially those of coherent correlation and adaptive optics. This assessment is based on the already demonstrated kilohertz frame rates of deformable mirror devices (DMD),[2] its small size (50 μm x 50 μm pixels) and its anticipated near unity diffraction efficiency. Since the micro-mechanical pixels are designed to move in a piston motion with more than a half wavelength of displacement,[3] reflected light may be phase modulated over a full 2π. This will allow exact phase matching of monochromatic illumination patterns up to the (spatial) Nyquist frequency $.5/\Delta_x$ where Δ_x is the spacing between adjacent pixels. The FBDMD is capable of approaching the 100 % efficient theoretical limit of the phase-only filter[4] as limited by its large area fill factor (currently 85 %) and its highly reflecting surface (currently evaporation-deposited aluminum.)

In practice the actual diffraction efficiency can be smaller for several reasons: 1) The microfabrication process for deformable mirror devices can leave unwanted material residues and residual strains that alter the hinge mechanical properties.[5] For the FBDMD this could cause pixel-to-pixel variation in displacement and unintended tilt; 2) Addressing circuitry variations, noise in electronics, and charge recombination all contribute to uncertainties about the absolute deflection of a deformable mirror pixel; 3) The constraints of a particular application effectively reduce the efficiency. A typical constraint that is also studied below is that of truncation errors caused by using binary phase-only filters (BPOF). Each of these effects leads to phase errors.

One practical approach to quantifying the effect of these errors on diffraction is to use statistical models in which the phase errors are considered to be random. We recently developed specific expressions for the expected value and standard deviation of the intensity of the far-field diffraction pattern statistical of an arbitrarily modulated phase-only spatial light modulator (SLM).[6] These are applicable to any phase-only SLM whose pixels are well modeled by piston and tilt. While the assumption that all random parameters describing the SLM modulation are statistically independent and identically distributed (i.i.d) is used to simplify the equations, experiments verifying these equations can be designed that ensure independence. For arbitrary modulation the equations may be efficiently calculated using a small number of fast Fourier transforms (FFT).

Even in devices where errors are negligible, applications can be envisioned in which pseudo-random modulation is used in order to maximize the information content represented by an SLM.[7] Evaluation of pseudo-random modulation of SLMs by deterministic simulation of diffraction is essentially a Monte Carlo procedure which provides only limited insight. Our equations are equally applicable to this situation, and may be extended to the case where random errors are present as well.

In this paper we consider the applicability of these equations to the analysis of phase-only SLMs by way of example numeric calculations and reductions of the equations to specific cases. In particular, we are interested in the performance of piston-only modulators such as the FBDMD, which may also be perturbed by random tilt errors. These equations are also either directly, or with slight modification, applicable to the performance of these SLMs in phase-only correlators. Before presenting these results we briefly summarize the equations describing expected value and standard deviation of the far field diffraction pattern. For details on their derivation consult ref. 6.

0-8194-0945-6/92/$4.00

2. STATISTICAL MOMENTS OF PHASE-ONLY SLMs

2.1 SLM Transmittance

The models for phase errors are described in Fig. 1. The pixels are treated as perfectly planar and unwarped but which may be tilted. The equations and subsequent plots are for reflective SLMs, and thus there is 4π of phase shift for one wavelength of physical displacement of the SLM surface. This leads to the model for the phase of the i'th pixel

$$\phi_i(x) = \frac{4\pi}{\lambda} [\ \psi_i' + \sin(\theta_i' x)\ ] \approx \psi_i + \theta_i x \tag{1}$$

in local coordinates. Since available DMDs have small tilts (θ_i' less than 10 °), the small angle approximation to the sine has also been used.

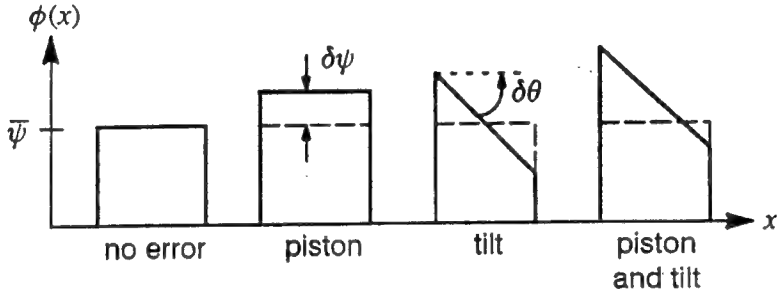


Fig. 1. The phase model for the SLM pixels.

The transmittance of an N-pixel, phase-only SLM is

$$t(x) = \sum_i r(x - x_i)\ \exp[\, j\phi_i(x - x_i)\,] \tag{2}$$

where r(x) = rect(x/w) is used to represent a pixel of width w. Any inactive area between the pixels has been treated as non-reflecting and the amplitude and phase of each pixel have been shifted to pixel locations x_i.

2.2 Expected SLM transmittance

The expected value of the transmittance of a phase-only SLM is

$$\bar{t}(x) = \langle t(x) \rangle = p^{\frac{1}{2}} \sum_i r(x - x_i)\ \exp[\, -\tfrac{1}{2}\sigma_\theta^2 (x - x_i)^2\,]\ \exp[\, j\bar{\psi}_i + \bar{\theta}_i (x - x_i)\,] \tag{3}$$

where the definition of the characteristic function[8] of the random variable u

$$\langle \exp(j\omega u) \rangle = \int_{-\infty}^{\infty} \exp(j\omega u)\, p_U(u)\ du = 2\pi\ \mathcal{F}^{-1}(p_U(u)) \tag{4}$$

has been used and where $\mathcal{F}$ is the Fourier transform operator. Eq. 3 is specific for tilt θ_i a gaussian random variable of standard deviation σ_θ. The probability distribution for piston ψ_i determines the constant p in Eq. 3. It is the square of the characteristic function of the piston error. For the case of gaussian distributed piston errors of standard deviation σ_ψ

$$p = \exp(-\sigma_\psi^2) \tag{5}$$

and for the case of uniformly distributed piston errors with total spread

$$a = \sqrt{12}\,\sigma_\psi \tag{6}$$

then

$$p = \mathrm{sinc}^2\left[\frac{a}{2\pi}\right] \tag{7}$$

2.3 Expected far-field intensity patterns

In general the expected intensity spectrum and the expectation of the squared intensity spectrum are

$$\bar{I}(f_x) = \langle T(f_x)\, T^*(f_x) \rangle = \mathcal{F}\{\langle t(x) \oplus t(x) \rangle\} \tag{8}$$

$$\langle I^2 \rangle = \bar{I}^2 + \sigma_I^2 = \langle TT^*\, T^* T \rangle = \mathcal{F}\{\langle t(x) \oplus t(x) \oplus t(x) \oplus t(x) \rangle\} \tag{9}$$

where $T(f_x)$ is mathematically the Fourier transform of t(x), physically the angular spectrum, and essentially the Fraunhofer diffraction pattern of the SLM. The correlation operator $\oplus$ indicates the integral

$$a(x) \oplus b(x) = \int a(x'+x)\, b^*(x')\, dx' \tag{10}$$

Eqs. 8,9 give the standard deviation of intensity $\sigma_I(f_x)$.

2.4 Expected far-field intensity patterns with piston and tilt error

Applying the i.i.d assumption for the parameters of tilt and piston, along with the linear phase model of eq. 1 to eqs. 8,9 leads to

$$\bar{I}(f_x) = \mathcal{F}\left\{ \bar{t}(x) \oplus \bar{t}(x) + [g_2(x) - p\, g_{11}(x)] \sum_i e^{j\bar{\theta}_i x} \right\} \tag{11}$$

$$\langle I^2(f_x) \rangle = 2\left[\bar{I}^2 - |\bar{T}|^4 \right] + \left| \bar{T}^2 + T_A \right|^2 + 4\, Re\left[\bar{T}\, T_B^* \right] + G_A$$

where

$$\begin{aligned} t_A(x) &= \sum_i \left[p^2 g_2(x - 2x_i) - p\, g_{11}(x - 2x_i) \right] \exp\left\{ j[\, 2\bar{\Psi}_i + \bar{\theta}_i (x - 2x_i)\,] \right\} \\ t_B(x) &= p^{\frac{1}{2}} \sum_i \left[g_3(x - x_i) - (2 + p^2)\, g_{21}(x - x_i) + 2p\, g_{11}(x - x_i) \right] \exp\left\{ j[\, \bar{\Psi}_i + \bar{\theta}_i (x - x_i)\,] \right\} \\ g_A(x) &= \left[g_4(x) + 4p\, g_{31}(x) - (2 + p^4)\, g_{22}(x) + 4p(2 + p^2)\, g_{211}(x) - 6p^2 g_{1111}(x) \right] \sum_i \exp(j\bar{\theta}_i x) \end{aligned} \tag{12}$$

where the g-functions are defined as

$$\begin{aligned} g_{ijkl}(x) &= g_i(x) \oplus g_j(x) \oplus g_k(x) \oplus g_l(x) \\ g_n(x) &= \left[r(x) \oplus \cdots \oplus r(x) \right]_n \exp\left(-\tfrac{1}{2} \sigma_0^2 x^2 \right) \end{aligned} \tag{13}$$

and n indicates an n-term correlation (n-1 integrals.)

2.5 Expected far-field intensity patterns with piston error

In the case of piston-only pixel modulation eqs. 11,12 further reduce to

$$\bar{I}(f_x) = |\bar{T}(f_x)|^2 + N q R^2(f_x) \tag{14}$$

$$\langle I^2(f_x) \rangle = 2\left[\bar{I}^2 - \tfrac{N-2}{N} |\bar{T}|^4 \right] - \tfrac{4}{N}\left[|\bar{T}|^2 + \tfrac{N}{2} q^2 R^2 \right]^2 + \left| \bar{T}^2 - p\, q\, T_A \right|^2$$

where

$$t_A(x) = \sum_i g_{11}(x - 2x_i) \exp(j2\bar{\Psi}_i) = r(x) \oplus r(x) * \sum_i \delta(x - 2x_i) \exp(j2\bar{\Psi}_i) \tag{15}$$

where "*" is the convolution operator and q = 1 - p. In all cases variables in the Fourier transformed variables are written as capitalized versions of their equivalent space domain variables.

3. COMMENTS ON THE FORM OF THE EQUATIONS

Eqs. 11,14 are seen to reduce to two terms. The first term in each is simply calculated as $|\mathcal{F}\{\langle t(x) \rangle\}|^2$, the magnitude squared of the Fourier transform of eq. 3. We will refer to this term as the intended SLM diffraction pattern or simply "signal". The name signal, however, presumes that the error-free phase is identical to the expected value of phase. Also, with reference to eq. 3, piston errors only attenuate the signal over the error free case. The tilt errors actually distort the signal by

convolving $T(f_x)$ (actually the Fourier transform of eq. 2 using the expected, rather than the random, values of piston and tilt) with the Fourier transform of a replicated array of truncated gaussian functions. This convolution or filtering process is responsible for aliasing of the signal unless $T(f_x)$ is bandlimited between the positive and negative Nyquist frequencies. The second term in eqs. 11 and 14 will be referred to as "pedestal". This term results from the Fourier transform correlation of two functions of pixel width w. This Fourier transform adds a broad base to the signal, and hence, the name pedestal. We think of it as the average level of phase error induced noise. We have chosen the term pedestal to distinguish it from the standard deviation of intensity, which is usually thought of as a measure of the "noise".

The two terms of signal and pedestal are apparent in Fig. 2. These curves represent the far-field diffraction pattern of a 128 pixel, 100 % fill-factor FBDMD which has all pixels nominally set to identical deflections and which is perturbed by various amounts of gaussian random piston errors. The curves are normalized so that the peak intensity in the error-free case is unity. The peak intensity decreases as the standard deviation of piston error increases. Between a standard deviation of .2 and .3 the coherent signal entirely disappears under the fully developed pedestal. The peak intensity of the pedestal is 1/128 that of the error-free case and its shape is described by $\text{sinc}^2(f_x)$, the intensity diffraction pattern of a single pixel. The zeroes of this element factor for a 100 % fill factor pixel are located at the diffraction orders of the grating/SLM.

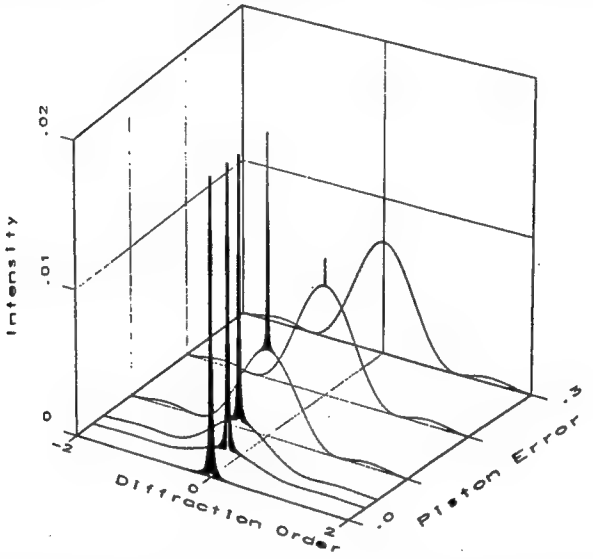


Fig. 2. Expected intensity of Fraunhofer diffraction for piston errors of standard deviation 0, .025, .05, .1, .2, .3 λ.

Fig. 3 is an expanded view around zero frequency of the curve in Fig. 2 with piston error of standard deviation .1 λ. The arbitrary unit, spot resolution, is 1/64 of a diffraction order. Fig. 3 also shows the $\pm\ \sigma_I(f_x)$ limits of the expected intensity and a single Monte Carlo run. The diffraction peak is reduced by a factor of .2 from the error-free case, but it is clearly distinguished from the sidelobes. Note that the $\langle I(f_x)\rangle - \sigma_I(f_x)$ curve is zero (not negative) over most of the sidelobe region. The Monte Carlo curve gives a qualitative feel that our statistical moment equations are correct, but numerous Monte Carlo runs must be averaged together to accurately estimate the mean and standard deviation of the diffraction pattern intensity.[6] In most applications the moments provide the desired information at less computational cost and are easier to interpret than Monte Carlo analysis.

Fig. 4 is a similar to Fig. 3 in that piston error is again .1 λ. In this case the simulation is of a diffraction grating that was designed by a simulated annealing procedure to generate 4 equal-intensity, equally spaced spots. The design procedure was similar to that of Dames[9] with the exception that it used eight phase levels rather than two. The diffraction pattern in Fig. 4 corresponds to illuminating 4 periods of the 32 pixel design. With the light shared between four diffraction peaks, and also approximation errors which reduce the diffraction efficiency by roughly 25 %, the reduced energy into the desired diffraction peaks leads to severe interference from the sidelobes.

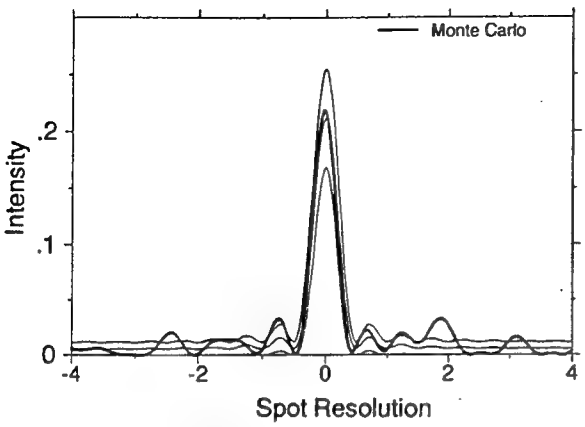


Fig. 3. Analysis of a nominally unmodulated phase grating with piston error of standard deviation .1 λ.

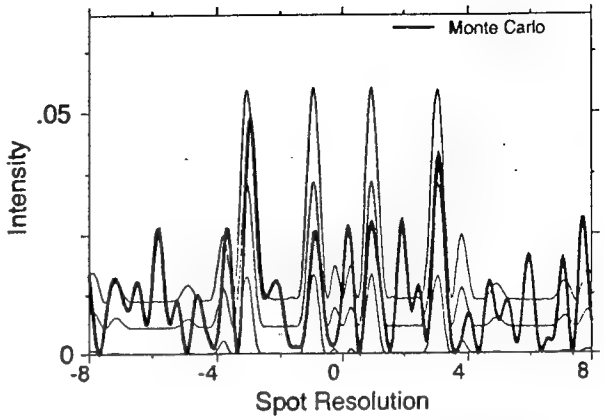


Fig. 4. Analysis of a four spot generating kinoform with piston error of standard deviation .1 λ.

Eqs. 11-14 are written in a form that is most easily programmed and calculated. In addition to the calculation of $\mathcal{F} < t(x) >$, the other space domain functions, which are of extent no wider than 4w, are computed by correlation. These basis functions are copied to multiple equations in a few arrays, multiplied by complex phasors and each of the arrays is Fourier transformed using the FFT. Most of the computation time is spent performing FFTs. Figs. 2-4 were calculated with 16,384 point FFTs with 9 FFT points per pixel.

4. DIFFRACTION EFFICIENCY OF PHASE-ONLY CORRELATORS

Figs. 2-4 may be viewed as the correlation peak resulting due to residual phase-matching errors in the filter plane of a correlator. Figs. 2-4 are analogous to responses of coherent optical correlators when the error-free target matches the reference template. In the case of a phase-only matched filter being illuminated by a phase-only signal spectrum Figs. 2-4 identically describe correlators. Fig. 4 represents the result for a filter that contains four references or a scene that contains four identical targets. With this interpretation, the peak intensities at $f_x = 0$ in Fig. 2 correspond to the effect of random piston errors on Horner efficiency.[10] Since the shape of the signal (i.e. correlation peak) is unchanged for piston errors (see Sec. 3) and since the phase-only correlator can have 100 % efficiency, the peak values of intensity are equivalent to Horner efficiency. These efficiencies are plotted in Fig. 5. Fig. 6 shows the reduction in peak intensity for various amounts of tilt. These may be related to Horner efficiency for the same reasons. In this case, the shape of the zero diffraction order can change with increasing tilt (see Sec. 3), but the change was insignificant over the range of tilts shown in the simulation.

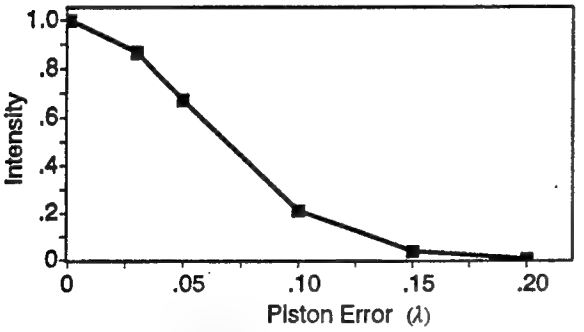


Fig. 5. Diffraction efficiency *vs.* piston error.

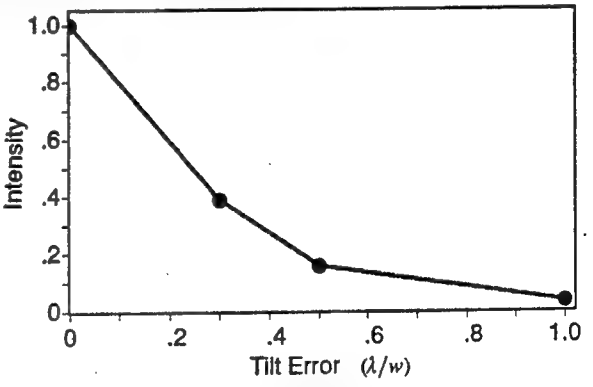


Fig. 6. Diffraction efficiency *vs.* tilt error.

4.1 Specific example: the binary phase-only filter

Fig. 5 is in fact the gaussian $p(\sigma_\psi)$ in eq. 5 as predicted by eqs. 3 and 11. An interesting point on this curve is piston error of standard deviation .072 λ. This corresponds to the binary phase-only matched filter (BPOF) in that this value of standard deviation is that for a uniform probability density function of total spread π. Diffraction efficiency for this case is 43.9 % . However, if the phase error is also modeled as uniformly distributed, which is more appropriate for truncation-type errors inherent in BPOF designs, then p would actually be a squared sinc function (eq. 6) and the diffraction efficiency would actually be 40.5 % . This may be compared with a simulation by Horner in which the reduction in diffraction efficiency of a BPOF relative to an analog phase-only filter 35.7 % and his calculation of efficiency for a deterministic square wave grating which was found to be 40.5 % .[11]

The relative error in the diffraction efficiency for random truncation errors in the design of a (tilt free) BPOF may also be found from eqs. 14,15. These equations can be simplified for the assumption of equally-spaced pixels of pitch Δ_x. Dividing the signal term of eq. 14 and the standard deviation of eq. 15 by the pedestal term in eq. 14 yields a signal-to-pedestal term

$$\frac{s}{n} = \frac{Np}{q}\, S^2(f_x) \tag{16}$$

and a standard deviation-to-pedestal term

$$\frac{\sigma_I}{n} = \sqrt{2\left[1 - \frac{2q}{N} - pS(2f_x)\right]\frac{s}{n} + p^2 S(2f_x) + 1 - \frac{q^2}{N}} \tag{17}$$

where

$$S(f_x) = \frac{sinc(N\Delta_x f_x)}{sinc(\Delta_x f_x)} \tag{18}$$

The element factor has been cancelled out which shows that these ratios are independent of pixel width w. Also these equations are periodic. Figs. 7,8 present one period of these equations (-.5 to .5 diffraction order) for a 16 pixel BPOF (uniform noise assumption). Fig. 7 shows the expected value of intensity (s/n + 1) and its one standard deviation bounds. It is most interesting that as the sidelobes become hidden in the pedestal the standard deviation approaches the value of pedestal. This is consistent with the statistics for the exponential probability distribution which are the same as those for fully developed laser speckle.[6,12] The signal-to-noise ratio (eq. 26 divided by eq. 27) is plotted in Fig. 8. The dashed line indicates unity signal-to-noise level. The relative error in diffraction efficiency is the inverse of Fig. 8 at zero frequency. For N much larger than 1/p the relative error at zero frequency is approximated as

$$\frac{\sigma_I}{s}(0) \approx q\sqrt{\frac{2}{Np}} \tag{19}$$

Even for 16 pixels the BPOF relative error (32.5 %) differs by less than 2 % with this approximation for relative error. For a 128 x 128 pixel FBDMD the relative error is around 1 %, assuming no other sources of error.

In passing we note that the relative error near the Nyquist frequency f_n (.5 diffraction order) is

$$\frac{\sigma_I}{s}(f_n) \approx \frac{N^2}{2}\sqrt{1+p^2}\ \frac{\sigma_I}{s}(0) \tag{20}$$

where we have used that the peak intensity of a sidelobe in this region is roughly $1/N^2$ of the intensity at zero frequency.

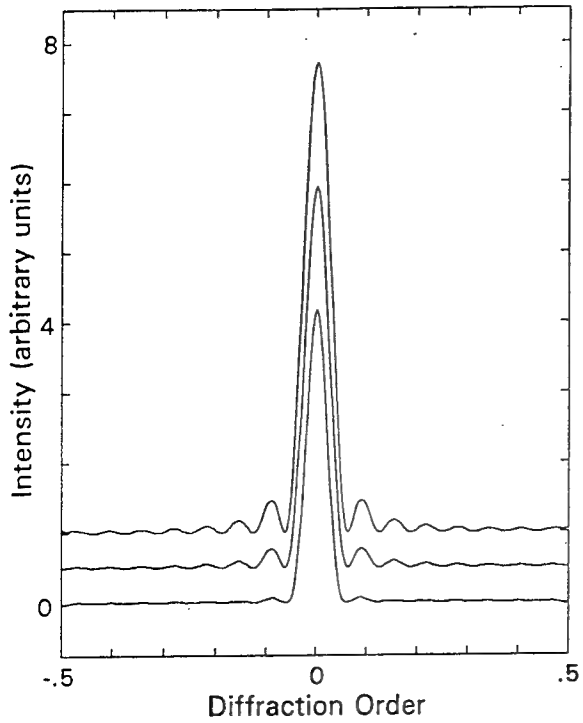


Fig. 7. Expected intensity and error bounds for 16 pixel binary phase-only filter. Curves are normalized by pedestal term $NqR^2(f_x)$.

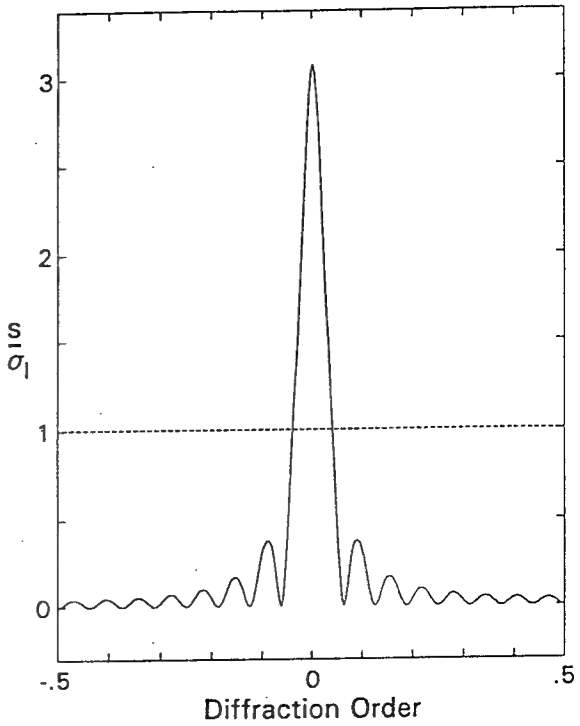


Fig. 8. Signal-to-noise ratio s/σ_I, which is the inverse of the relative error.

5. EXTENSIONS TO THE MODELS

Sec. 4 considered the diffraction pattern of a nominally unmodulated FBDMD to be representative of the correlation surface of a phase-only correlator when the reference and target are the same. Equations can be developed that explicitly consider both the SLM transmittance t(x) and the SLM illumination u(x) containing the spectrum of the target. The assumption of statistical independence of pixel transmittances $a_i(x) = r(x)\exp[\phi_i(x)]$ (expressed in local coordinates) and non-random illumination leads to the intensity diffraction pattern

$$< |C(f_x)|^2 > = \mathcal{F} < u(x)t(x) \oplus u(x)t(x) > \tag{21}$$

$$= \mathcal{F}\left\{u(x)\bar{t}(x) \oplus u(x)\bar{t}(x) + \sum_i \int u_i(x'+x)\left[< a_i(x'+x)\,a_i^*(x') > - \bar{a}_i(x'+x)\bar{a}_i^*(x')\right]u_i^*(x')\,dx'\right\}$$

where $u_i(x) = u(x + x_i)$ has been used in order to express rectangularly sampled line segments of u(x) and a(x) in local coordinates.

5.1 Expected correlation intensity for arbitrary non-random SLM illumination

For the case of pixel phase described in terms of i.i.d. piston and gaussian i.i.d. tilt errors this simplifies to

$$< |C(f_x)|^2 > = \mathcal{F}\left\{u(x)\bar{t}(x) \oplus u(x)\bar{t}(x)\right. \tag{22}$$

$$\left. + \sum_i \left([\,u_i(x)r(x) \oplus u_i(x)r(x)\,]\exp(-\tfrac{1}{2}\sigma_\theta^2 x^2) - p\,[\,u_i(x)r(x)\exp(-\tfrac{1}{2}\sigma_\theta^2 x^2) \oplus u_i(x)r(x)\exp(-\tfrac{1}{2}\sigma_\theta^2 x^2)\,] \right)\right\}$$

The correlation surface $C(f_x)$, which is usually thought of as in the spatial coordinates, has been expressed in frequency coordinates to be consistent with the equations presented above.

5.2 Expected correlation intensity for bandlimited non-random SLM illumination

If u(x) is sufficiently bandlimited so that it may be approximated in terms of its samples $u_i(0)$, eq. 22 approximates to

$$< |C(f_x)|^2 > \approx \mathcal{F}\left\{ u(x)\bar{t}(x) \oplus u(x)\bar{t}(x) + [\, g_2(x) - p\, g_{11}(x)\,] \sum_i |u_i(0)|^2 \exp(j\bar{\theta}_i x) \right\} \quad (23)$$

This result is closely similar to eq. 11 with the illumination weighting the signal and pedestal terms. This equation could then be used to calculate Horner efficiency for amplitude weighted illumination in a similar manner to Sec. 4.1.

5.3 Expected correlation intensity for random phase-only SLM illumination

For the case that u(x) is a phase-only signal with random phase modulation that is uncorrelated with and has identical statistics as the phase of t(x) then the further expectation of the u(x) terms in eq. 22 gives

$$< |C(f_x)|^2 > = \mathcal{F}\left\{ \bar{t}(x)\bar{t}^{*\prime}(x) \oplus \bar{t}(x)\bar{t}^{*\prime}(x) + 2[\, g_2(x) - p\, g_{11}(x)\,] \exp(-\tfrac{1}{2}\sigma_\theta^2 x^2) \sum_i \exp[\, j(\bar{\theta}_i - \bar{\theta}_i')x\,] \right\} \quad (24)$$

where $t^{*\prime}(x) = u(x)$ has been used in order to emphasize the correlation being sought. As was true eq. 23, eq. 24 also is quite similar to eq. 11. In this case the pedestal term is scaled by a factor of 2 [one due to t(x) and one due to t´(x).] An especially important change is that the influence of errors is the square of that in eq. 11. For example, assume there are only piston errors, then the signal is attenuated by p^2 rather than p. Furthermore if t(x) and t´(x) are completely random ($a = 2\pi$) and uncorrelated, say pseudo-random sequences, then p=0 and the pedestal term completely describes the expected intensity of the optical cross correlation.

5.4 Correlated phase models

To this point the random variables have been assumed to be independent. There are however many practical situations in which the random variables are correlated. Slow spatial variations in a microfabrication process (say for fabricating DMDs) might also cause slowly varying changes in pixel deflection across a chip. White gaussian noise added to video signals that electrically-control SLMs can become correlated (colored) if the signal is filtered; and filtering is frequently unavoidable: A/D and D/A conversion (e.g.frame grabbers), resampling an image to a different resolution, and signal conditioning all require filters. In DMDs and other SLMs using CCDs in the addressing circuitry, both non-unity charge transfer efficiency (though in practice usually high enough to ignore[13]) and the sampling of continuous signals, filters and thus correlates the noise. Therefore we expect to see correlated pixel errors in many cases.

Correlated errors do not necessarily invalidate the theory presented to this point. Not only can similar theories be developed for correlated models, but also experiments can be designed that minimize correlation. The simplest model for the correlated pixel errors is to assume that they arise from sampling a continuous signal. The signal noise is modeled as a stationary stochastic process derived by passing gaussian white noise through a linear filter. For lowpass filtering of gaussian noise we expect to see additional bandlimiting of the pedestal. In experiments, if the control signal is filtered, the specular signal will be filtered, as well. For special cases, such as correlations over a few pixels, it is practical to develop expressions similar to eq. 11. For instance, if only adjacent pixels are correlated eq. 14 contains two additional integrals that are associated with halo. These integrals represent impulses of width 2w, centered at $\pm\Delta x$ for equi-spaced pixels. The Fourier transform of these two integrals together with the integral in eq. 14 represents the pedestal in the case of nearest neighbor correlated pixels. Extension to any correlation distance is always possible though the expressions become unwieldy.

In experiments, depending on the source of the noise, it is also possible to observe the test signal and whiten its noise. The deterministic portion of the signal would, in the process, also have its high frequencies enhanced. An ideal solution for experiments is to generate white noise using a thermal noise source or pseudo-random code generator. If the signal is brick-wall filtered at the Nyquist frequency of a sampler or A\D convertor then the samples are known to be independent.[14] These samples can be computer generated off-line by any number of computer routines and written into a video frame grabber or programmable waveform generator. Thus, it appears likely that we will be able to closely approximate in experiments the inherent assumptions of eqs. 11-15.

6. CONCLUSIONS

We have analyzed the effects of random piston and tilt errors on the diffraction patterns of phase-only spatial light modulators. In general we observe that piston and especially tilt error must be well controlled in phase-only SLMs in order to perform coherent optical processing with large diffraction efficiency and low noise. The analysis also describes the operation of correlators which use phase-only SLMs in filter plane. One special case is the binary phase-only matched filter correlator which by this analysis is 40.5 % as efficient as the phase-only correlator, and is in agreement with earlier deterministic simulations and modeling. The analysis equations presented here are especially useful when one is concerned about rapidly simulating the effect of phase errors on correlation with very rich and textured images.

7. ACKNOWLEDGEMENTS

This work was supported in part by DARPA, through RADC, on contract F19628-92-K-0021.

8. REFERENCES

1. L.J. Hornbeck, "Deformable-mirror spatial light modulators," *Proc. SPIE* 1150, 86-102 (1989).
2. R.M. Boysel, "A 128 x 128 frame-addressed deformable mirror spatial light modulator," *Opt. Eng.*, 30, 9, 1422-27 (Sept 1991).
3. R.M. Boysel, J.M. Florence, and W.R. Wu, "Deformable mirror light modulators for image processing," *Proc. SPIE* 1151, 183-194 (1989).
4. J.L. Horner, "Phase-only matched filtering," *Appl. Opt.*, 23, 6, 812-16 (15 March 1984).
5. W.R. Wu, R.O. Gale, L.J. Hornbeck, and J.B. Sampsell, "Electro-optical performance of an improved deformable mirror device," *Proc. SPIE* 825, 24-31 (1987).
6. R.W. Cohn and R.J. Nonnenkamp, "Statistical moments of the transmittance of phase-only spatial light modulators," *Proc. SPIE*, V. 1751-31 (1992).
7. See Shannon's coding theorem, for instance in P. LaFrance, *Fundamental Concepts in Communication*, Prentice Hall, Englewood Cliffs, NJ (1990).
8. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York (1965).
9. M.P. Dames, R.J. Dowling, P. McKee, and D. Wood, "Efficient optical elements to generate intensity weighted spot arrays: design and fabrication," *Appl. Opt.*, 30, 19, 2685-91 (1991).
10. J.L. Horner, "Light utilization in optical correlators," *Appl. Opt.*, 21, 24, 4511-14 (15 December 1982).
11. J.L. Horner and J.R. Leger, "Pattern recognition with binary phase-only filters," *Appl. Opt.*, 25, 5, 609-11 (1 March 1985).
12. *Laser Speckle and Related Phenomena*, second edition, J.C. Dainty, ed., Springer, Berlin (1984).
13. E.L. Dereniak and D.G. Crowe, *Optical Radiation Detectors*, Ch. 9, Wiley, New York 1984.
14. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York (1965).

Limited phase modulation and its effect on phase-only correlation

Robert W. Cohn

University of Louisville, Department of Electrical Engineering
Louisville, Kentucky 40292

Joseph L. Horner

Rome Laboratory, Optical Signal Processing Branch
Hanscom Air Force Base, Massachusetts 01731

ABSTRACT

Phase modulating devices, especially spatial light modulators, are often incapable of producing a full 360 degrees of phase modulation. Other limitations due to calibration errors, signal distortion and quantization can cause the actual phase modulation to differ from the desired modulation. Such limitations on the filter plane modulator can reduce the performance of phase-only correlators. We quantify these performance losses for various phase limitations, both through simulation and through the development of an approximate model of performance. In one case we quantitatively compare the performance of phase-only filters that are optimized for limited-range phase modulation (as prescribed by Juday's "minimum Euclidean distance" principle) with the performance of non-optimal filters. In another case we analyze the effect of not compensating a quadratic dependence of phase on signal voltage which is anticipated for some spatial modulators, e.g. deformable mirror devices.

1. INTRODUCTION

Ideally, phase-only correlation is performed with a filter plane spatial light modulator (SLM) that produces the conjugate phase spectrum of the object that is to be recognized. However, current SLMs usually cannot produce all desired values of phase, and thus, the limited range of modulation can lead to a reduction in performance. Various phase limitations are shown in Fig. 1. In each case the desired value of phase ϕ_d is mapped to an achievable value (the actual) value of phase ϕ_a. We view the SLM as producing a prescribed amount of phase mismatch (or systematic error) $\delta\phi = \phi_d - \phi_a$ from the ideal phase-only filter. For each value of phase desired there is a known, or systematic, amount of phase error according to the known phase mapping characteristic (e.g. Fig. 1.). The best known example of limited phase modulation is where the SLM produces less than a full 2π modulation range (Fig. 1a., solid line.) There are numerous potential sources for these various limitations. In general, the SLM might be physically incapable of producing the desired modulation, or alternatively, the signal to be placed on the modulator could be distorted. For instance, analog SLMs addressed by quantized signals can be limited in exactly the same way as an SLM that only produces quantized levels. This point of view allows one to consider, in addition to the intrinsic performance of the SLM, the combined performance of the SLM along with its associated address electronics.

The recognition of limited range SLMs has led to filter design procedures that maximize various performance metrics under the constraint of limited modulation range.[1-7] Juday has shown that optimal performance can be obtained for any specified SLM limitation by a mapping between the optimal filter (for which the modulation can be perfectly general) and a filter that is achievable with the SLM.[3] The optimal mapping minimizes the mean squared error (i.e. the "Euclidean distance") between the desired filter and the achievable filter. Using this approach, after a filter is designed, the metric for the optimal achievable filter can be directly calculated and compared with all other filters (achievable or otherwise). However the value of the metric depends strongly on the exact image to be recognized in each design -- thus a new simulation is required for each new image. This approach works well for prescribing an optimized design, but it can become cumbersome, especially in that numerous simulations can be required, such as if one wishes to perform a detailed sensitivity study with respect to a disparate set of images and various SLM constraints. We have instead developed an approximate model of the effect of limited range phase modulation on correlation. The model provides new insight into the operation of phase only correlation, it is easy to calculate and, while approximate, it permits quantitative evaluations of various phase limitations.

2. MODEL OF PEAK CORRELATION AMPLITUDE

The new model follows from observations that the correlation peaks can be viewed as a summation of a large number of coherent wavefronts. This summation is similar in form to an expected value or ensemble average. This correspondence with statistical models provides insight into correlation and is used in the development of our model of correlator performance. The correspondence with statistical models is seen as follows: In the phase-only correlator the signal s(x) is Fourier transformed

0-8194-1541-3/94/$6.00

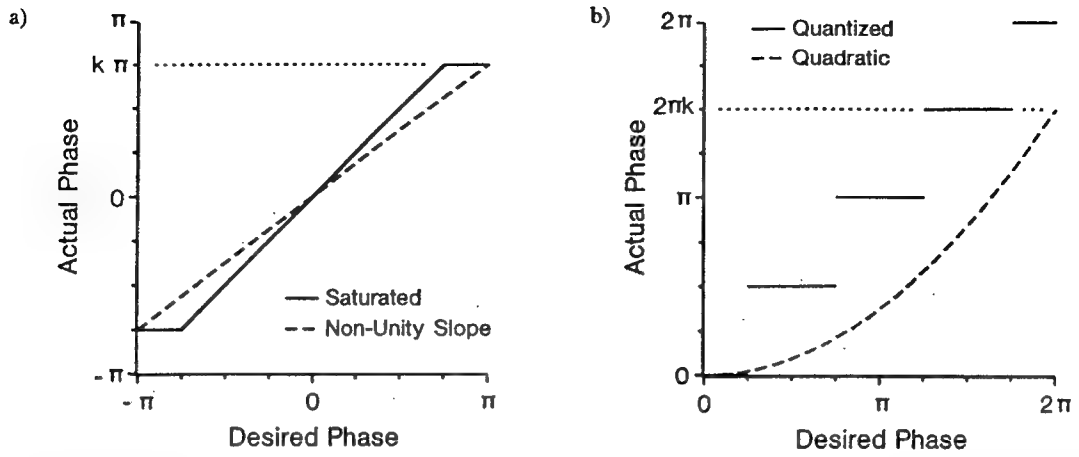


Fig. 1. Phase limitations shown as mappings from desired to actual phase. a) Continuous mappings for which actual phase has less than 2π range of modulation. b) Mappings for m=4 levels of quantization with k=(m-1)/m and quadratic distortion.

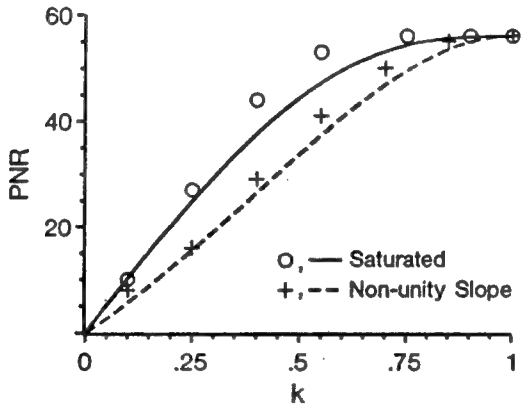


Fig. 2. Comparisons of PNR model (curves) with PNR simulation (data points) for the phase limitations described in Fig. 1a. The image used for the input and filter is the face of a woman.

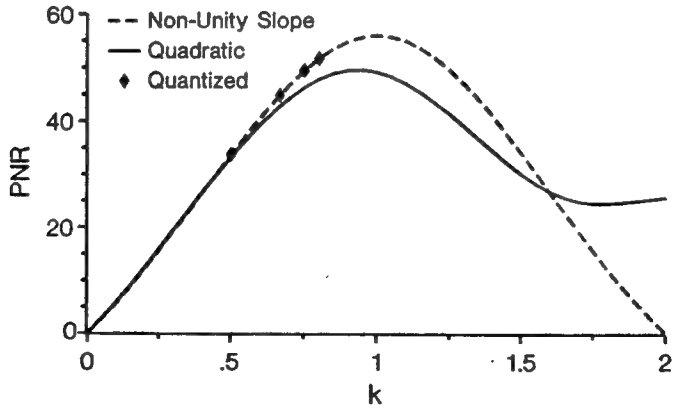


Fig. 3. PNR of non-unity slope mapping compared with the quadratic and quantized mappings (described in Fig. 1b.)

to produce

$$S(f) = a(f)\exp[j\phi(f)] = \mathcal{F}\{s(x)\} \tag{1}$$

where a(f) is the positive valued amplitude spectrum and ϕ(f) is the phase spectrum. The phase-only, like the matched filter, correlator is designed so that the filter exactly cancels all phases. This produces a peak correlation plane amplitude on the optical axis of

$$c(0) = \int_{-B_f/2}^{B_f/2} a(f)\, df = B_f\, \bar{a} \tag{2}$$

where the overline is used to indicate the spatial average (across the frequency plane) of the variable a. This average is identical in form to the temporal average.[8,9] If a limited-range phase-only spatial light modulator is used in the filter plane then the peak correlation amplitude becomes

$$c(0) = \int_{-B_f/2}^{B_f/2} a(f)\exp\{j\,\delta\phi[\phi(f);k]\}\, df = B_f\, \overline{a\exp(j\,\delta\phi)} \tag{3}$$

where the phase error $\delta\phi$ is a function of ϕ and depends on the parameter k. For instance, the simplest expression for this is $\delta\phi = (1\text{-}k)\phi$ for the non-unity slope curve (fig. 1a.) Eq. 4 can be rewritten as

$$c(0) = B_f \int_0^{\infty}\int_{-\pi}^{\pi} a\exp(j\delta\phi)\, p(a,\phi)\, d\phi\, da = B_f\, \langle a\exp(j\delta\phi)\rangle \tag{4}$$

Thus the spatial average in eq. 3 is equivalent to the integral in eq. 4. It takes the form of and approximates an ensemble average or expected value of the random variable contained between the brackets < > (i.e. the ensemble average operator.) The density function p(a,ϕ) that arises when performing a transformation of variables in integration represents a histogram of the ordered pairs (a,ϕ). As such, it approximates the joint probability density function of the random variable (a,ϕ). This correspondence between deterministic and random quantities is no different than intensity histogramming which is commonly used in image processing; except that we are using complex-valued rather than real occurrences.

In order to appreciate how limited phase affects correlation we make the following simplifying approximations to eq. 3

$$c(0) \approx B_f\, \bar{a}\, \overline{\exp(j\delta\phi)} \approx \frac{B_f\, \bar{a}}{2\pi}\int_{-\pi}^{\pi}\exp(j\delta\phi)\, d\phi \tag{5}$$

A heuristic explanation for the first approximation is that phase more strongly influences correlation than amplitude. This approximation will generally be better if the values of phase error do not depend on the values of amplitude and is exactly true if a and ϕ are independent random variables. The second approximation is true if the phase ϕ is uniformly distributed over 2π. This assumption is definitely false for reference objects which are real-valued, symmetric and centered on the optical axis. In this case phase takes on only values of 0 and π and a binary phase only filter will produce identical performance as the full range phase-only filter. Also if the phase of the image spectrum is non-uniformly distributed and the modulator phase is limited, there is a phase offset that optimizes correlation (the threshold line angle.)[3-7] As far as our analysis is concerned there is no preferred phase offset under the assumption of uniformly distributed phase. However, also note that we have made histograms of images from commonly available digital image libraries and found that many do have phase that is roughly uniform in distribution.

Specific functional expressions for peak correlation amplitude with the phase limitations shown in Fig. 1 are found by evaluating eq. 5. These are

$$c_s(0) = B_f\, \bar{a}\,[k + (1-k)\mathrm{sinc}(1-k)] \tag{6}$$

$$c_n(0) = B_f\, \bar{a}\,\mathrm{sinc}(1-k) \tag{7}$$

$$c_q(0) = B_f\, \bar{a}\,\mathrm{sinc}\left(\frac{1}{m}\right) = c_n(0)\Big|_{k=\frac{m-1}{m}} \tag{8}$$

where the correlation amplitude for the saturated characteristic is c_s; the non-unity slope is c_n; and the quantized case is c_q. The second equality in eq. 8 indicates that using k=(m-1)/m where m is the number of quantization levels, that correlation

performance is identical for both the non-unity slope characteristic and the quantized characteristic. Quadratic phase limitation has been evaluated numerically and is presented in a subsequent graph.

3. MODEL OF PEAK-TO-NOISE RATIO

The effect of systematic phase errors only acts through the peak amplitude, for many of the recently discussed correlation metrics.[11,12] One of these metrics is the peak-to-noise ratio (PNR)

$$PNR = \frac{|c(0)|}{\sqrt{\frac{1}{B_x - \Delta_x}\left[\int_{-B_x/2}^{B_x/2} |c(x)|^2\,dx - \int_{-\Delta_x/2}^{\Delta_x/2} |c(x)|^2\,dx\right]}} \tag{9}$$

where the denominator represents the RMS amplitude across a correlation plane of spatial bandwidth B_x which excludes a small region of width Δ_x centered around the correlation peak. For an SLM of N pixels of width Δ_x ($B_x = N\Delta_x$) and using eq. 5 and properties of the Fourier transform eq. 9 can be approximated as

$$PNR \approx \frac{\sqrt{N-1}\;|\,\bar{a}\,\overline{\exp(j\delta\phi)}\,|}{\sqrt{\overline{a^2} - |\,\bar{a}\,\overline{\exp(j\delta\phi)}\,|^2}} = \sqrt{\frac{N-1}{Z\;\overline{\exp(j\delta\phi)}^{-2} - 1}} \tag{10}$$

where the parameter

$$Z = \frac{\overline{a^2}}{\bar{a}^2} \tag{11}$$

This model for PNR, eq. 10, depends on only three parameters, N the number of pixels in the frequency plane SLM, k the parameter which controls the amount of phase error for a given type of phase limitation, and Z which solely describes the distribution of the amplitude spectrum. It is especially useful that our model only depends on the image data to this small degree and that quite varied images can produce identical values of Z (as can be seen by considering eq. 11.) Using our statistical interpretation (correspondence between the spatial average and the ensemble average) it can be seen that the numerator of Z is simply a variance of the spectral amplitudes. Minimizing the variance (or standard deviation) minimizes Z. A white spectrum (over the full width of the phase-only SLM) will produce the smallest value of Z, which is unity and consequently it will produce the largest values of PNR. In the case of no phase errors PNR is infinite indicating that all energy has been focused to a single pixel in the correlation plane and no energy coincides with the other pixels. Infinite PNR is not physically possible if the frequency plane illumination is of finite extent. Nonetheless it is obtained as a result of designing our model to match the discretely sampled images used for our FFT-based simulations (described below.)

4. COMPARISONS OF PHASE LIMITATIONS ON PEAK-TO-NOISE RATIO

In the simulations, the phase-only filters are either the conjugate phase spectra of the image to be recognized (for k=1) or the conjugate phase with the phase limitations as indicated in Fig. 1. No additional phase offset was added to any filter (i.e. we did not perform optimization with respect to the threshold line angle.) For the examples shown here the image used is a woman's face.[12] It is a 64x64 array of pixels and is padded by zeroes in a 128x128 array. The correlation is performed by fast Fourier transforming the image and multiplying its spectrum by a 128x128 array of unit magnitude, complex-valued numbers which represent the N=16,384 pixels of a phase-only SLM. The value of Z for this image is 6.23 and this value is used for Z in calculations of eq. 10.

4.1. Simulations compared with model

According to Juday's theory, even if the line angle is not adjusted, the saturated phase mapping produces a larger PNR than the non-unity gain mapping (both shown in Fig. 1a.)[3] This is relationship is evident for the simulated PNRs in Fig. 2 (with "+" representing PNR for non-unity slope and "o" representing PNR for the saturated phase mapping.) The curves represent the model for the two types of phase errors. Specifically eq. 6 (for non-unity slope) or eq. 7 (for saturation) is substituted into eq. 10. We find that the curves for this and other images favorably track the simulated results for the purposes of making relative comparisons between the effects of various systematic phase errors. The fact that the simulated PNR exceeds the modeled PNR is an indication that more of the image spectrum can be phase matched by the limited-phase SLM than for a uniformly distributed spectrum, for this particular image. Other images (or line angles) can produce simulated PNR that is less than the modeled PNR. We do note that for each value of k and for either model or simulation, the saturated phase mapping

produces greater PNR than the linear mapping; as expected according to Juday's theory. Uniquely in contrast to Juday's approach is that a simple model (only requiring the vaguest of information about the image) has been used to quantify the difference in performance for the two different types of phase limitations.

4.2. Comparisons of linear, quadratic and quantized phase mappings

These results are plotted in Fig. 3. The linear (non-unity slope) is the same curve as shown in Fig. 2, except that k ranges from zero to two. Simulations of performance for the quadratic mapping (Fig. 1b) were presented previously in ref. 13, though for a different metric. Note that the value of k=1 corresponds to $\phi_m = 2\pi$ in ref. 13. Additionally the performance for 2 to 5 levels of phase quantization are plotted in Fig. 3. As indicated by eq. 8, PNR for the quantized mappings is equivalent to that for non-unity slope at k=.5, .67, .75 and .8. The quadratic mapping is practically indistinguishable from the non-unity slope mapping for k < .5. These values are all noticeably less than the PNR for the saturated mapping (shown in Fig. 2.) For k=1 the non-unity slope and the saturated mappings produce no phase mismatch and thus they produce PNR identical with the ideal phase-only filter. It may at first appear surprising that the PNR for the quadratic mapping with k approximately equal to 1 produces PNR greater than 90 % of that for the ideal filter. However, the phase mismatch for any value of phase never exceeds one quarter wavelength (90° maximum at $\phi_d = \pi$.) While this may appear to be a large amount of phase error, there is an analogous result for imaging systems: They are considered to behave as essentially diffraction-limited if the phase errors are less than a quarter wavelength.[14] PNR that is nearly identical to that for the maximum PNR for the quadratic case for 4 levels of quantization (k=.75.) Again we note that the maximum phase error is one-quarter wavelength.

5. CONCLUSIONS

These findings indicate that several correlation metrics of interest can be modeled with reasonable accuracy by employing the approximation of eq. 5. The model reduces to easily computed functions for phase errors of current interest which can arise from limitations in current SLMs and tradeoffs between cost and performance of electronic interface circuits. Since the model requires only general knowledge of image properties (namely the bandwidth parameter Z, which is identical for any number of different images,) it relieves system designers from the necessity of performing image-based simulations every time a design parameter, component specification or tolerance changes.

6. ACKNOWLEDGEMENTS

RWC acknowledges the sponsorship of the Advanced Research Projects Agency (ARPA) through U.S. Air Force Rome Laboratory contract F19628-92-K-0021.

7. REFERENCES

1. M.W. Farn and J.W. Goodman, "Optimal maximum correlation filter for arbitrarily constrained devices," *Appl. Opt.* 28, 3362-3366. (15 August 1989)
2. R.D. Juday, "Correlation with a spatial light modulator having phase and amplitude cross coupling," *Appl. Opt.*, 28, 22, 4865-4869. (15 November 1989)
3. R.D. Juday, "Optimum real filter and the minimum Euclidian distance principle," *Appl. Opt.*, 32, 26, 5100-5111. (10 September 1993)
4. M.W. Farn and J.W. Goodman, "Optimal binary phase-only matched filters," *Appl. Opt.*, 27, 4431-4437. (1988)
5. M.W. Farn and J.W. Goodman, "Bounds on the performance of continuous and quantized phase-only matched filters," *JOSA A*, 7, 1, 66-72. (1 January 1990)
6. B.V.K.V. Kumar, "Tutorial survey of composite filter designs for optical correlators," *Appl. Opt.*, 31, 32, 4773-4801. (10 August 1992)
7. B.V.K.V. Kumar and Z. Bahri, "Efficient algorithm for designing a ternary valued filter yielding maximum signal to noise ratio," *Appl. Opt.*, 28, 10, 1919-1925. (15 May 1989).
8. A. Yariv, *Optical Electronics*, 4th ed. Chs. 1 and 10. Saunders College Publishing, Philadelphia. (1991)
9. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 1st ed. McGraw-Hill, New York. (1965)
10. J.L. Horner, "Metrics for assessing pattern-recognition performance," Appl. Opt., 31, 2, 165-166. (10 January 1992)
11. B.V.K.V. Kumar and L.G. Hassebrook, "Performance measures for correlation filters," *Appl. Opt.*, 29, 20, 2997-3006. (10 July 1990)
12. University of Southern California Image processing Institute Image Data Base. The woman's face is image 5.1.4.
13. J.L. Horner and P.D. Gianino, "Effects of quadratic phase distortion on correlator performance," *Appl. Opt.*, 31, 20, 3876-3878. (10 July 1992)
14. W.J. Smith, *Modern Optical Engineering*, Ch. 11. McGraw-Hill, New York. (1966)

a reprint from Applied Optics

Effects of systematic phase errors on phase-only correlation

Robert W. Cohn and Joseph L. Horner

The performance of phase-only optical correlators is usually reduced if the filter-plane phase differs from that prescribed for the classical matched filter. Current spatial light modulators, which frequently produce less than 2π phase modulation, and interface circuits, which quantize or incorrectly amplify signals placed on the spatial light modulator, both can produce systematic phase errors. We examine these effects using a model of correlation-peak amplitude as a function of phase error. The correlation peak is reasonably approximated as the product of an average of unity-amplitude error phasors multiplied by the average amplitude across the filter plane. The trends predicted by this new model compare favorably with computer simulations that use gray-scale images.

Key words: Phase-only correlators, pattern recognition, spatial light modulators, correlation metrics, coherent optical processors, real-time correlators, binary optics, optimal filters.

Introduction

Limitations of Spatial Light Modulators Used for the Filter Plane

The phase-only correlator consists of an amplitude-only spatial light modulator (SLM) in the input plane and a phase-only transparency in the filter plane.[1] The phase of the filter is, by definition, the negative of the phase spectrum of the object to be recognized. The enthusiasm for the phase-only correlator has centered in great part around its higher diffraction efficiency, its narrower correlation peaks in relation to those for matched-filter correlators, and its ease of writing on available SLM's. However, it has been difficult to realize SLM's that are continuously variable over a 2π range of phase modulation and that cause no residual modulation of amplitude.[2,3] The recognition of these practical limitations is probably a major reason for the current popularity of the binary phase-only filter in optical correlators. Also, in recognition of the problem of limited modulation range of SLM's, several studies have been devoted to developing filter design procedures that maximize various performance metrics under the constraint of limited modulation range.[2–8] The emphasis of the optimal filter design studies is more on minimizing the effects of limited modulation range than on evaluating and understanding the effects on performance. Most similar among these studies in relation to the current study is that of Farn and Goodman,[6] especially in their developing a lower performance bound for correlation with quantized phase-only filters that are optimized by adjustments of a phase offset with respect to the threshold line angle.[5,8] We note below an interesting correspondence between one of Farn and Goodman's results and our's.

Instead, we choose to focus directly on describing the effects of limited-phase modulation characteristics on correlation. We view limited-phase SLM's as producing a phase error $\delta\phi$, which is the difference between ϕ_d, the phase required for (ideally, the 2π range) phase-only correlation and ϕ_a, the phase actually achievable with a constrained SLM. The comparisons presented here make no special provisions for optimizing filter performance in the presence of a limited-phase constraint, especially since our most basic model does not use detailed information about the object to be recognized. However, a more general model is discussed that shows how the spectral distribution of the object with respect to the threshold line angle influences correlator performance.

In general, phase errors arise from unintended transformations of signals to be placed on the SLM. Thus phase errors can also result from the nonideal performance of electronic and optical components

R. W. Cohn is with the Department of Electrical Engineering, University of Louisville, Louisville, Kentucky 40292; J. L. Horner is with Rome Laboratory, Hanscom Air Force Base, Massachusetts 01731-5000.

Received 30 July 1993; revised manuscript received 23 February 1994.

0003-6935/94/235432-08$06.00/0.

between the signal source and the SLM. Therefore this study is applicable to many practical issues that must be considered in the design and the development of actual phase-only correlators. Some possible sources of phase error are suggested in the next subsection, but in general the practical issues relate to calibration of signal levels and to the tolerances and the accuracy of components that process the signal that is ultimately transformed into the phase modulation.

In this paper we are interested specifically in systematic as opposed to random errors. By systematic phase error we mean that the phase-error distribution $\delta\phi(f)$ across the filter plane (of frequency coordinate f) is nonrandom or deterministic. Recent studies on effects of systematic phase errors have focused primarily on simulation and experimental measurement[9,10] rather than on developing models of correlator performance, as is done in this study. The study by Downie *et al.*[6] included the effects of a static pattern of phase errors across the SLM, whereas the study by Horner and Gianino[9] considered cases in which phase error at each pixel is describable as an explicit function of the phase desired. While both static and filter-dependent phase errors are types of systematic error, our study considers specifically the filter-dependent case.

Our proposed model follows from observations that the correlation peaks can be viewed as a summation of a large number of coherent wave fronts. This summation is similar in form to an expected value or an ensemble average. This correspondence with statistical models provides insight into correlation and is used in the development of a simple analytic model of correlator performance. The expressions that describe performance of phase-only correlators as a function of systematic phase-error characteristics are derived below and are shown to have good correspondence with computer simulations.

Phase-Error Characteristics of Interest

Figure 1 illustrates a variety of systematic errors that can be reasonably anticipated in nonideal phase-only SLM's and practical optoelectronic correlation systems. Figure 1(a) shows two types of continuous phase errors. Both curves are typical of SLM's that produce phases only between $\pm k\pi$. The solid curve represents an SLM that matches exactly the signal phase within these limits. The dashed curve has a slope of value k, and thus phase error $\delta\phi$ is proportional to $1 - k$. The solid transfer characteristic could result from either the limited modulation range of the SLM or the saturation of amplifiers in the signal interface circuits. The dashed curve could result from an incorrect amplifier gain setting, and, as drawn, the saturated SLM could also produce the nonunity slope characteristic if its signals were attenuated by the factor k.

Figure 1(b) shows a discontinuous phase transfer curve. This is a generalized characteristic for the binary phase-only filter, in which the two phases may

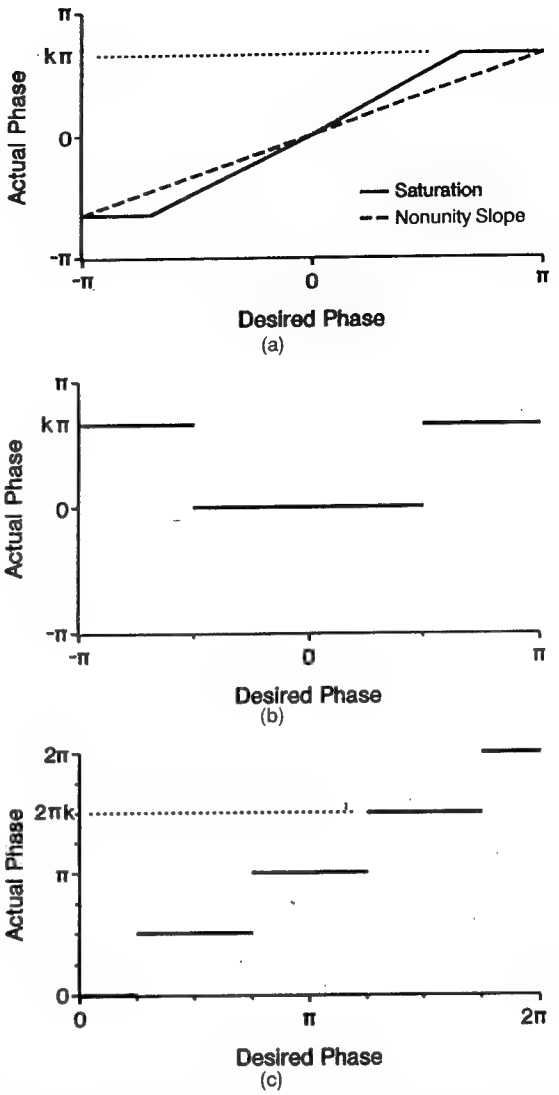


Fig. 1. Various systematic phase errors: (a) two continuous mappings for which the actual phase has less than a 2π range of modulation, (b) binary phase-only mapping in which the range of phase modulation is not necessarily π, (c) mapping in which the actual phase is quantized into m levels [in which Eq. (1) gives the relationship between k and m].

be separated by other than π. Such an error can result from incorrect fabrication of a binary optic or, once again, from incorrect gain of the signal applied to the SLM. Another type of discontinuous phase characteristic is that resulting from quantization of the phase, as illustrated in Fig. 1(c). Quantization is to be expected for filters fabricated by binary optical techniques or for SLM's that are addressed by digital circuits. For this case, k is related to the number of quantization levels, m, as

$$k = \frac{m - 1}{m}. \tag{1}$$

For example, Fig. 1(c) shows four phase levels, and thus k equals 0.75. This relationship is used [specifically, in Eq. (12)] to map the performance curves for quantization errors onto the performance curves for slope errors. In all subsequent analyses, k is always assumed to be less than or equal to one.

A secondary objective of this paper is to compare specifically the effects on performance of the two continuous phase-modulation characteristics shown in Fig. 1(a). It was our trying to understand these differences that first suggested the analysis presented in this paper. We present this comparison as a detailed demonstration and confirmation of our theory. As further background on this problem, we now review what is already known about the effect of these SLM phase errors and how the current design-oriented methodologies could also be used to understand their effects.

Juday has shown that, if the filter-plane SLM produces less than a 2π range of modulation, then the optimal performance under this constraint is obtained by mapping of the optimal phase under no constraints [or in Fig. 1(a), the desired phase] to the achievable (or actual) phase according to the saturated curve.[4] Furthermore, there is a value of phase offset (adjusted with respect to the threshold line angle[5,8]) that minimizes the mean-squared error (or, in Juday's terminology, the Euclidean distance) between the optimal filter and the filters achievable with the SLM. With Juday's approach, after a filter is designed to optimize a specific performance metric, the metric for the optimal achievable filter can be directly calculated and compared with all other filters (achievable or otherwise). However, the value of the metric depends strongly on the exact image to be recognized in each design; thus a new simulation is required for each new image. Juday's approach can be generalized directly to incorporate systematic phase errors. One can do this by altering systematically the phase-modulation curve by the known error and then by redesigning the filter. This approach is perfect for design optimization but becomes cumbersome, requiring numerous simulations, for systems analysis of the dependence of correlation on disparate imagery and SLM phase errors. For this reason we instead choose to develop a simple model that, while sacrificing the numerical accuracy of available approaches, provides general insight into correlator performance. Also, note a difference in comparing our results to the earlier research on optimized phase-only filters[2–8]: In our derivation we make simplifying assumptions about the image spectrum that permit the dependence on threshold line angle to be removed altogether.

Model of Peak-Correlation Amplitude

The performance of optical correlators is characterized by any one of several metrics.[11,12] These usually depend strongly, and for several metrics, exclusively, on the peak amplitude of the correlation surface. Our analyses can then be understood by recognition of the similarity between the peak-correlation amplitude and a specific type of average across the filter plane.

In the phase-only correlator the signal $s(x)$ is Fourier transformed to produce

$$S(f) = a(f)\exp[j\phi(f)] = \mathcal{F}[s(x)], \tag{2}$$

where $a(f)$ is the positive-valued spectral amplitude and $\phi(f)$ is the spectral phase. The phase-only correlator, similar to the matched-filter correlator, is designed so that the filter exactly cancels all phases. This produces a peak-correlation plane amplitude on the optical axis of

$$c(0) = \int_{-B_f/2}^{B_f/2} a(f)\mathrm{d}f = B_f\bar{a} \tag{3}$$

for an SLM of finite extent or spatial bandwidth B_f. The overbar in Eq. (3) is used to indicate the spatial average (across the frequency plane) of the variable a, where the spatial average for a general function $g(f)$ and for a given bandwidth B is defined as

$$\bar{g} \equiv \frac{1}{B}\int_{-B/2}^{B/2} g(f)\mathrm{d}f. \tag{4}$$

This form is identical to the temporal average.[13,14] If the SLM exhibits systematic phase errors, then the peak-correlation amplitude becomes

$$\begin{aligned} c(0) &= \int_{-B_f/2}^{B_f/2} a(f)\exp\{j\delta\phi[\phi(f); \mathrm{k}]\}\mathrm{d}f \\ &= B_f\overline{a\exp(j\delta\phi)}, \end{aligned} \tag{5}$$

where phase error $\delta\phi$ is a function of ϕ and depends on parameter k. For instance, the simplest expression for this is $\delta\phi = (1 - k)\phi$ for the nonunity slope curve [Fig. 1(a)].

A transformation of variables also permits Eq. (5) to be written as

$$\begin{aligned} c(0) &= B_f\int_0^{\infty}\int_{-\pi}^{\pi} a\exp(j\delta\phi)p(\phi, a)\mathrm{d}\phi\mathrm{d}a \\ &= B_f\langle a\exp(j\delta\phi)\rangle. \end{aligned} \tag{6}$$

Thus the spatial average in Eq. (5) is equivalent to the integral in Eq. (6), which takes the form of an ensemble average or an expected value, defined by

$$\langle g\rangle \equiv \int_{-\infty}^{\infty} gp(g)\mathrm{d}g, \tag{7}$$

where $p(g)$ is the probability-density function of the random variable g. Even though a and ϕ are known quantities and are not thought of as random variables, the collection of these variables across the

frequency plane permits them to be mathematically evaluated in the same way as random variables. This correspondence provides useful insights into our modeling method. Thus $p(a, \phi)$ is the joint density of occurrences of the pair of values a and ϕ. A histogram of the ordered pairs (a, ϕ) approximates the joint-density function.

For our purposes of developing a general understanding of how systematic phase errors affect correlation we make the following simplifying approximations to Eq. (6):

$$c(0) \approx B_f \bar{a}\, \overline{\exp(j\delta\phi)} \approx \frac{B_f \bar{a}}{2\pi} \int_{-\pi}^{\pi} \exp(j\delta\phi)\mathrm{d}\phi. \quad (8)$$

A heuristic explanation for the first approximation is that phase influences correlation more strongly than amplitude. This approximation is generally better if the values of phase error do not depend on the values of amplitude. The second approximation is true if the phase ϕ is uniformly distributed over 2π. This assumption is definitely false for reference objects that are real valued, symmetric, and centered on the optical axis. In this case, phase takes on values of 0 and π only. We made histograms of images from commonly available digital-image libraries and found that many do have phases that are rougly uniform in their distribution.

In general, phase does depend on amplitude, and the correspondence between Eq. (5) and approximation (8) is most clearly seen by rewriting of Eq. (6) as

$$c(0) = B_f \int_0^{\infty} ap(a)\mathrm{d}a \int_{-\pi}^{\pi} \exp(j\delta\phi)p(\phi|a)\mathrm{d}\phi, \quad (9)$$

where the standard identity from statistics $p(a, \phi) = p(\phi|a)p(a)$ is used and $p(\phi|a)$ is the density of ϕ conditioned on the value of the parameter a. For cases in which $p(\phi|a)$ does not depend on a, then a and ϕ are by definition independent, in which case the first approximation in approximation (8) is equivalent to Eq. (9). It also not necessary that a and ϕ be independent. It is sufficient that the arguments of the ensemble average are uncorrelated; i.e., $\langle a \exp(j\delta\phi)\rangle = \langle a\rangle \langle \exp(j\delta\phi)\rangle$. Equation (9) can thus be used to resolve differences between our conceptual models based on approximation (8) and experimental results obtained by images with known joint distributions of amplitude and phase. An example in which the form of the joint distribution is an issue is that of designing binary phase-only and ternary filters. It is exactly the distribution of amplitude and phase that determines the optimal choice of phase offset with respect to the threshold line angle.

Correlation-Peak Amplitude as a Function of the Systematic Phase Errors

The expressions for peak-correlation amplitude found by evaluation of approximation (8) with the specific functions of ϕ given in Fig. 1 are

$$c_s(0) = B_f \bar{a}[k + (1 - k)\mathrm{sinc}(1 - k)], \quad (10)$$

$$c_n(0) = B_f \bar{a}\ \mathrm{sinc}(1 - k), \quad (11)$$

$$c_q(0) = B_f \bar{a}\ \mathrm{sinc}\left(\frac{1}{m}\right) = c_n(0)|_{k=m-1/m}, \quad (12)$$

$$|c_b(0)| = \frac{B_f \bar{a}}{\pi/2} \cos\left[\frac{\pi}{2}(1 - k)\right] = \cos\left[\frac{\pi}{2}(1 - k)\right] c_q(0)|_{m=2}, \quad (13)$$

where the correlation amplitude for the saturated characteristic is c_s, the nonunity slope is c_n, the binary phase is c_b, and the quantized case is c_q. As indicated by the second equality in Eq. (12), which uses the definition from Eq. (1), the correlation performance as a function of k is identical for both the nonunity slope characteristic and the quantized characteristic. It is interesting to note that we have derived the same result as Eq. (12) by assuming that phase-mismatch errors from quantization are uniformly distributed random variables of spread $\pm m/2$.[15] In Eq. (13), absolute value signs are included to eliminate a phase-offset term that is not of interest. For k equal to unity the correlation amplitude for the binary characteristic is identical to the two-level quantized characteristic, as expected. Equation (13) shows that the binary phase-only correlator is fairly insensitive to phase offsets that are not exactly π. Also, the amplitude of the binary phase-only correlator (for $k = 1$) is identical to the nonunity slope correlator when $k = 0.5$.

Another systematic error of recent interest is the quadratic phase error that can arise as a result of the quadratic relationship between address voltage of a deformable-mirror pixel and the phase modulation it produces.[9] Approximation (8) can be applied in the same manner as before. The resulting expression contains several Fresnel integrals.

An additional observation is in order on the functional form of Eq. (12). This term appears in previous studies on quantized phase-only filters[6] and kinoforms.[16,17] Of special interest is Farn and Goodman's results on the lower bound of performance for correlators using quantized phase-only filters [their Eq. (39)]. This ratio of signal-to-noise ratio when phase is quantized to signal-to-noise ratio when there is no quantization is identical to the ratio of $c(0)$ in Eq. (12) to $c(0)$ when there is no quantization. It is interesting to note that Farn and Goodman's performance bound is applicable to filters that are optimized under the constraint of quantized phase (i.e., on the basis of the threshold line angle), while under our set of assumptions of uniformly distributed phase and independent amplitude, optimization is not required and does not improve the performance of our model. Therefore our equality for nonoptimized quantized filters coincides with Farn and Goodman's lower bound. Perhaps this result can be used in the

establishment of performance bounds with other types of phase constraint.

Models of Correlation Metrics: Continuous and Discrete Signals

As mentioned above, the effect of systematic phase errors acts only through the peak ampltide for many of the recently discussed correlation metrics.[11,12] We show the dependence on peak amplitude for one of these metrics and also make clear the correspondence between our continuous-signal model and the discrete-signal models that are typically used to simulate optical correlator performance.

Peak-to-noise ratio (PNR) is

$$\mathrm{PNR} = \frac{|c(0)|}{\left\{\frac{1}{B_x - \Delta_x}\left[\int_{-B_x/2}^{B_x/2} |c(x)|^2 \mathrm{d}x - \int_{-\Delta_x/2}^{\Delta_x/2} |c(x)|^2 \mathrm{d}x\right]\right\}^{1/2}}, \tag{14}$$

where the denominator represents the rms amplitude across a correlation plane of spatial bandwidth B_x that excludes a small region of width Δ_x. This metric is the continuous-signal equivalent to peak-to-correlation energy (PCE″) for discrete signals:

$$\mathrm{PCE}'' = \frac{(N-1)^{1/2}\,|c_0|}{\left(\sum_i |c_i|^2 - |c_0|^2\right)^{1/2}}, \tag{15}$$

where $N = B_x/\Delta_x$ is the number of sample points in an image.[11] (The metric PCE″, some of its properties, and its relationship to Kumar and Hassebrook's PCE[12] is described in Ref. 11.) The equivalence between PNR and PCE″ follows by insertion of the standard piecewise approximation for integrals into Eq. (14):

$$g_i \equiv \frac{1}{\Delta_x}\int_{-\Delta_x/2}^{\Delta_x/2} g(x + i\Delta_x)\mathrm{d}x \approx g(i\Delta_x), \tag{16}$$

which is valid as long as the function varies slowly enough over the limits of integration Δ_x. In the case of digital simulations of optical correlation (including those simulations presented here), which typically use the fast-Fourier transform (FFT), B_x represents the extent of the FFT window and Δ_x represents the separation between FFT sample points. For these analyses we also assume that the SLM in the frequency plane has N pixels of pitch Δ_f that extends over a bandwidth B_f. The signal $s(x)$ is also spatially limited to $B_x/2$ or, equivalently, s_i is limited to $N/2$ samples so that the SLM can sample the signal spectrum $S(f)$ at or below the half-Nyquist rate. For these sets of assumptions, PNR is approximately equal to PCE″ to within the limits set by the Nyquist sampling theorem.[18]

For these assumptions, the PNR in Eq. (14) can be rewritten as

$$\mathrm{PNR} = \frac{(N-1)^{1/2}|c(0)|}{[E_c - |c(0)|^2]^{1/2}}, \tag{17}$$

where, through applying Parseval's relation to the phase-only correlator, we can express the correlation energy E_c as

$$E_c = \int_{-B_x/2}^{B_x/2} |c(x)|^2 \mathrm{d}x = \int_{-B_f/2}^{B_f/2} |a(f)\exp(j\delta\phi)|^2 \mathrm{d}f$$

$$= \int_{-B_f/2}^{B_f/2} a^2(f)\mathrm{d}f. \tag{18}$$

Thus, with reference to Eq. (5), it can now be seen that the dependence of PNR on $\delta\phi$, or, equivalently, the parameter k, is caused by the peak amplitude $c(0)$ only. Explicitly, substituting identity (4), approximation (8), and Eq. (18) into Eq. (17), our model for PNR becomes

$$\mathrm{PNR} \approx \frac{(N-1)^{1/2}|\bar{a}\,\overline{\exp(j\delta\phi)}|}{[\overline{a^2} - |\bar{a}\,\overline{\exp(j\delta\phi)}|^2]^{1/2}}, \tag{19}$$

where the approximate equality is due to approximation (8). Except for this approximation, which we are evaluating in this paper, approximation (19) is equivalent to Eq. (15). For modeling purposes it is most convenient to view the averaging operator as a continuous integral, while for simulation with signals represented by their spectra the averages are well approximated as

$$\bar{g} \approx \frac{1}{N}\sum_{i=1}^{N} g_i, \tag{20}$$

which follows directly from the validity of approximation (16).

Discussion on the Form of Peak-to-Noise Ratio

Our model for PNR, approximation (19), depends on only three parameters: N, the number of pixels in the frequency-plane SLM; k, the parameter that controls the amount of phase error for a given type of systematic error characteristic; parameter

$$Z = \frac{\overline{a^2}}{\bar{a}^2}, \tag{21}$$

which depends solely on the distribution of the values of the spectra. Using our statistical interpretation (correspondence between the spatial average and the ensemble average), we can see that the numerator of Z is simply a variance. Minimizing the standard deviation of the spectral values minimizes Z.

In its most compact form, approximation (19) is then written as

$$\mathrm{PNR}^2 = \frac{(N-1)|\overline{\exp(j\delta\phi)}|^2}{Z - |\exp(j\delta\phi)|^2}. \tag{22}$$

Thus in the phase-only correlator the value of PNR is also influenced by the spectral distribution of the signal that is to be recognized. A white spectrum (over the full width of the phase-only SLM) will produce the smallest value of Z, which is unity, and consequently gives the largest values of PNR. In the case of no phase errors, PNR is infinite, indicating that all energy is focused to a single pixel in the correlation plane and no energy coincides with the other pixels. Infinite PNR is not physically possible if the frequency-plane illumination is of finite extent, but it is nonetheless obtained in simulations as a result of designing our model to match the discrete model used for FFT-based simulations. The infinite value of PNR is a direct result of the discrete Fourier transform of the white spectrum producing the (discrete-signal) delta function.[11]

An example showing the sensitivity of PNR to spectral distribution for a reasonably broad set of model spectra is presented. Since we usually observe low-pass spectra when working with real-world images, we choose a variable-exponent power law:

$$a(f) = \left(1 - \left|\frac{f}{w}\right|^p\right)\mathrm{rect}\left(\frac{f}{2w}\right), \tag{23}$$

where $2w$ represents the spectral width of the signal and p is assumed to be greater than zero. The resulting value of the bandwidth parameter Z is

$$Z = \frac{p+1}{2p+1}\frac{B_f}{w}. \tag{24}$$

The term $(p + 1)/(2p + 1)$ only varies between 0.5 and 1 for all values of p between 0 and ∞.

This analysis can be extended directly to two-dimensional rectangularly separable spectra with the result that

$$Z = Z_{f_x} Z_{f_y}, \tag{25}$$

where Eq. (24) is specified independently for the f_x and f_y coordinates. For two-dimensional circularly symmetric spectra with radial power-law dependence identical to Eq. (23),

$$Z = \frac{p+2}{p+1}\frac{B_f^2}{\pi w^2}. \tag{26}$$

For $p = \infty$ (the spectrum is essentially a circ function) and $2w = B_f$, Z in Eq. (26) is $4/\pi$. The minimum value of Z is greater than unity for the circular spectrum, but this is to be expected since it has a smaller total bandwidth than the rectangular spectra. Basically, Eqs. (24)–(26) indicate that there is an effective bandwidth for every spectral distribution that produces identical correlator performance. For understanding PNR, the equating of Z values from different distributions determines the definition of effective bandwidth.

Comparison of Modeled and Simulated Peak-to-Noise Ratio

In the simulations the phase-only filters are either the conjugate-phase spectra of the image to be recognized (for $k = 1$) or the conjugate phase with the systematic phase errors, as indicated by Fig. 1(a). In none of the simulations did we add phase offsets to the filter (in order to optimize performance; see discussions above on threshold line angle). The images consist of a 64 × 64 array of pixels padded by zeros in a 128 × 128 array. We performed the correlation by fast-Fourier transforming the image and multiplying its spectrum by a 128 × 128 array of unit-magnitude complex-valued numbers that represent the $N = 16{,}384$ pixels of the phase-only SLM.

Figure 2 shows phase-only correlator performance for the saturated and nonunity slope phase-error characteristics for the specific input image of a woman's face.[19] The curves represent Eq. (21) for the two types of phase errors. We used a value of $Z = 6.23$ in the model in order to produce the same value of PNR as was simulated for no phase errors (i.e., for $k = 1$). This is not a free parameter but rather a recognition that for no phase error, approximation (8) (our principal approximation) is equivalent to Eqs. (3) and (5). We find that the curves for this and other images track favorably the simulated results for the purposes of making relative comparisons between the effects of various systematic phase errors. The fact that the simulated PNR exceeds the modeled PNR is an indication that more of the image spectrum can be phase matched by the limited-phase SLM than for a uniformly distributed spectrum, for this particular image. Other images (or phase offsets) can produce a simulated PNR that is less than the modeled PNR. The finer resolution of these effects, along the lines of

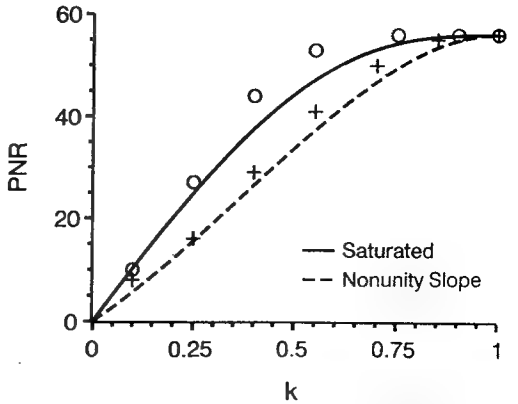


Fig. 2. Comparisons of PNR model (curves) with PNR simulation (data points) for the systematic errors described in Fig. 1(a). The image used for the input and the filter is the face of a woman.[19]

Eq. (9), is probably of practical interest only in sophisticated pattern-recognition problems for which the image data sets have rather specialized properties.

We also observed the same trends, in fact, with even closer agreement, for the phase-only correlation of a tank image[19] with the same tank on a natural cluttered background (Fig. 3). In this case we know that the undesired signal produces additional phase errors and alters the spectral distribution of amplitude. The value of Z is again set to force the model and simulation to be equivalent if there are no systematic phase errors. However, in this case, Z incorporates these additional affects as well. For clutter modeled as a noise process the additional phase errors can be analyzed by the approaches described in Refs. 7 and 20. The closer agreement between the model and the simulation in this case is probably because of the additional randomization of phase that is caused by the addition of clutter.

The simulation with the image of the woman (and images of smaller bandwidth) show that the PNR depends essentially on peak amplitude $c(0)$. For its specific value of $Z = 6.23$, ignoring variations in the denominator of approximation (19) changes the values of the PNR curve only by ~8%.

Since spectral bandwidth is used widely, we consider briefly the range of bandwidths for which the denominator of approximation (19) has as small an effect on PNR as does the test image. We do this by determining the relative bandwidth $2w/B_f$ for each spectrum that gives a value of Z equivalent to that of the test image. For brick-wall spectra (i.e., power $p = \infty$) the one-dimensional model has a relative bandwidth of 16%, the square-separable model [Eqs. (24) and (25)] has a relative bandwidth of 40%, and the circular model [Eq. (26)] has a relative bandwidth of 44%. For spectra that are most heavily weighted toward dc (i.e., $p = 0$) the relative bandwidth becomes

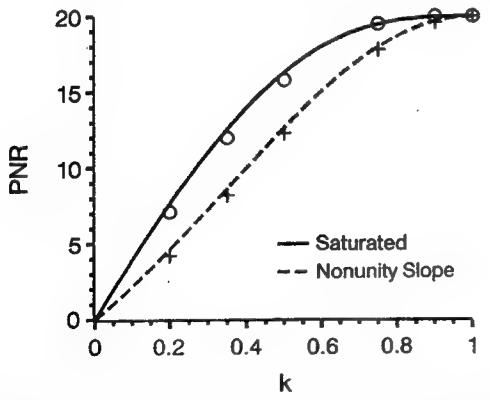


Fig. 3. Comparisons of PNR model (curves) with PNR simulation (data points) for the systematic errors described in Fig. 1(a). The image used for the filter is a tank. The image used for the input is the same tank with clutter added.[19] The model does not include the effect of the clutter. However, for purposes of comparison the value of the model at the end point ($k = 1$) is set equal to that for the simulation.

32% for one-dimensional, 80% for square-separable, and 64% for circular spectra. While the bandwidths vary for the brick-wall-model spectra between 16% and 44%, the effective bandwidth area (for example, the rectangular case $B = B_x B_y = 16\%$), or equivalently, the effective number of SLM pixels illuminated is constant (2621 for a 16,384-pixel SLM).

In summary of this section, we examined the effects on correlation when the filter-plane SLM is incapable of modulation over the full 2π range. We performed a detailed comparison of a phase-only filter correlator for which the desired phase is mapped to a limited-phase SLM in two different ways: linear scaling of phase and saturation of phase. For each value of k and for either model or simulation, the saturated phase mapping produced a greater PNR than the linear mapping, as expected according to Juday's theory. Uniquely in contrast to Juday's approach is that a simple model (requiring only the vaguest of information about the image) was used to quantify the difference in performance for the two different phase characteristics.

Conclusions

These findings indicate that several correlation metrics of interest can be modeled with reasonable accuracy by use of approximation (8). The model reduces to easily computed functions for phase errors of current interest that can arise from limitations in current SLM's and trade-offs between cost and performance of electronic interface circuits. Since the model requires only general knowledge of image properties (namely, the bandwidth parameter Z, which is identical for any number of different images), it relieves system designers from the necessity of performing image-based simulations every time a design parameter, a component specification, or a tolerance changes.

This is not the first evidence that correlation metrics for a variety of images show the same trends. Even though phase errors were not considered, Kumar and Hassebrook found similar trends for simulations of various correlation metrics with a variety of images.[12] It appears that their results can also be explained in terms of an effective image bandwidth. The most interesting and surprising result of this study as observed by Kumar and Hassebrook and ourselves is the very limited dependence of correlation performance on the detailed parameters defining an image. Such results are analogous to nonparametric statistics[21] in which the resulting statistic (the correlation metric) is, at best, only weakly dependent of the functional form of the underlying distribution (the image).

We thank R. D. Juday for helpful technical discussions on generalized optimality criteria for constrained correlation filters, physical insights into correlation metrics, and some clarifying examples of both. R. W. Cohn acknowledges the sponsorship of the Advanced Research Projects Agency through U.S.

Air Force Rome Laboratory contract F19628-92-K-0021.

References and Notes

1. J. L. Horner and P. D. Gianino, "Phase-only matched filtering," Appl. Opt. **23,** 812–816 (1984).
2. M. W. Farn and J. W. Goodman, "Optimal maximum correlation filter for arbitrarily constrained devices," Appl. Opt. **28,** 3362–3366 (1989).
3. R. D. Juday, "Correlation with a spatial light modulator having phase and amplitude cross coupling," Appl. Opt. **28,** 4865–4869 (1989).
4. R. D. Juday, "Optimum realizable filters and the minimum Euclidean distance principle," Appl. Opt. **32,** 5100–5111 (1993).
5. M. W. Farn and J. W. Goodman, "Optimal binary phase-only matched filters," Appl. Opt. **27,** 4431–4437 (1988).
6. M. W. Farn and J. W. Goodman, "Bounds on the performance of continuous and quantized phase-only matched filters," J. Opt. Soc. Am. A **7,** 66–72 (1990).
7. B. V. K. Vijaya Kumar, "Tutorial survey of composite filter designs for optical correlators," Appl. Opt. **31,** 4773–4801 (1992).
8. B. V. K. Vijaya Kumar and Z. Bahri, "Efficient algorithm for designing a ternary-valued filter yielding maximum signal-to-noise ratio," Appl. Opt. **28,** 1919–1925 (1989).
9. J. L. Horner and P. D. Gianino, "Effects of quadratic phase distortion on correlator performance," Appl. Opt. **31,** 3876–3878 (1992).
10. J. D. Downie, B. P. Hine, and M. B. Reid, "Effects and correction of magneto-optic spatial light modulator phase errors in an optical correlator," Appl. Opt. **31,** 636–643 (1992).
11. J. L. Horner, "Metrics for assessing pattern-recognition performance," Appl. Opt. **31,** 165–166 (1992).
12. B. V. K. Vijaya Kumar and L. G. Hassebrook, "Performance measures for correlation filters," Appl. Opt. **29,** 2997–3006 (1990).
13. A. Yariv, *Optical Electronics,* 4th ed. (Saunders, Philadelphia, Pa., 1991), pp. 3, 356–359.
14. A. Papoulis, *Probability, Random Variables, and Stochastic Processes,* 1st ed. (McGraw-Hill, New York, 1965), pp. 323–332.
15. R. W. Cohn, "Random phase errors and pseudorandom phase modulation of deformable mirror spatial light modulators," in *Optical Information Processing Systems and Architectures IV,* B. Javidi, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **1772,** 360–368 (1992).
16. J. W. Goodman and A. M. Silvestri, "Some effects of Fourier-domain phase quantization," IBM J. Res. Dev. **14,** 478–484 (1969).
17. W. J. Dallas, "Phase-quantization: a compact derivation," Appl. Opt. **10,** 673–676 (1971).
18. A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, N.J., 1975), pp. 26–30.
19. University of Southern California Image Processing Institute Image Data Base. The woman's face is image 5.1.4, and the tank is image 7.1.7.
20. R. W. Cohn and R. J. Nonnenkamp, "Statistical moments of the transmittance of phase-only spatial light modulators," in *Miniature and Micro-Optics: Fabrication,* C. Raychoudhuri and W. B. Veldkamp, eds., Proc. Soc. Photo-Opt. Instrum. Eng. **1751,** 289–297 (1992).
21. W. W. Daniel, *Applied Nonparametric Statistics.* 2nd ed. (PWS Kent, Boston, Mass., 1990), pp. 18–21.

Performance models of correlators with random and systematic phase errors

Robert W. Cohn

University of Louisville, Department of Electrical Engineering
Louisville, Kentucky 40292

Joseph L. Horner

Rome Laboratory, Optical Signal Processing Branch
Hanscom Air Force Base, Massachusetts 01731

ABSTRACT

Hybrid correlators are composed of numerous non-ideal electronic and optical components that, to one degree or another, limit performance through unintended transformations of signals. Many of these effects show up as phase errors at a spatial light modulator (SLM) plane. The errors can be described as random variables, or as systematic offsets from the correct phases, as appropriate. Sources of systematic phase errors include quantizing circuits, incorrect or nonlinear amplifier gain, limited range phase modulators and residual phase modulation of amplitude-mostly SLMs. Random phase errors arise from electronic noise and fabrication variations of SLMs. In this paper several systematic and random filter plane errors are related through a single parameter that describes the amount of phase mismatch. A model of peak-to-noise ratio (PNR) is also presented that describes the combined effects of random and systematic errors. This expression contains the products of two functions, one that depends only on systematic, the other on random, phase mismatch. PNR is also a function of the number of pixels in the filter plane modulator and a normalized moment the amplitude of the image spectrum. The model is useful for developing phase error budgets for correlation systems.

1. INTRODUCTION

Optical correlation is highly dependent on the amount of phase mismatch or wavefront error across the filter plane. If the filter perfectly conjugates or matches the incident wavefront then essentially all the energy is focused into a single bright spot at the correlation plane. If the filter poorly matches the incident wave then the filter behaves like a diffuser in that the correlation intensity is greatly reduced and that a broad background of scattered light is produced akin to a speckle pattern.[1] It has been noted for various types of phase errors that small to moderate errors, say less than a quarter wave, produce slight losses in correlation intensity and add only slight amounts of background noise to the correlation plane.[2,3] A quite similar situation is that imaging systems are known to produce, essentially, diffraction-limited resolution if the total aberrations are less than one quarter wavelength.[4] As in imaging systems, the critical issue is the total amount of phase error rather than the functional description of the phase error. The physical reason for this is that the formation of a correlation peak is due to the coherent superposition of a large number of coherent wavefronts. Each cell or pixel in the filter plane contributes a wavefront with a specific amount of phase error and thus, the amount of filter plane phase error, rather than its spatial distribution, appears to be the critical factor.

We have recently begun modeling the coherent formation of the correlation peak as a statistical process. The model is applicable even when the phase errors are non-random (i.e. systematic). Our earlier work on phase errors considered systematic[5] and random[6-8] phase errors individually. In this paper we now are able to make direct comparisons between the relative effects of various systematic and random phase errors. The model can also be used to calculate the combined effects of systematic and random errors. The earlier work also focused on phase-only filters. The models here apply to any fully-complex filter (including the special case of the phase-only filter).

2. BACKGROUND: DIFFRACTION FROM THE FILTER PLANE TO THE CORRELATION PLANE

Classes of phase error. We consider phase error to be any deviation from a planar wavefront. In our current work on optical correlation we have been evaluating how residual phase errors on the field transmitted through the filter influence the formation of correlation peaks. In this situation the phase error can be defined as the difference between ϕ the phase of the signal spectrum and ϕ_f the phase of the filter. The phase error expressed as a function of f the position across the filter plane is

$$\delta\phi(f) = \phi(f) - \phi_f(f) \tag{1}$$

0-8194-1544-X/94/$6.00

If the errors are random they can be described by a probability distribution. We will generally assume that the random portion of the phase errors are unbiased with respect to a plane wavefront. If there is a bias term, we consider this to be the systematic portion of the phase error. One broad class of systematic phase errors is filter-dependent or SLM-limited errors.[5] These phase errors can be described as an explicit, non-random function of the values of the signal phase. For example, one type of systematic phase error is a gain or slope error in which the phase error

$$\delta\phi(f;k) = (1-k)\,\phi(f) \tag{2}$$

is linearly proportional to the value of the signal phase and where the parameter k indicates the degree of phase match. Other examples are shown in Table 1, that is discussed in Sec. 4.

General results from our previous work on the effects of phase errors are reviewed in the remainder of this section. The equations from ref. 8, originally described Fourier diffraction between any two Fourier planes. Those expressions are rewritten here in terms of filter plane and correlation plane fields. The expressions from ref. 5 on systematic phase errors are identical; however, we show that that analysis for phase-only filters also applies to fractional power filters[9] if the spectral amplitude of the signal is replaced with the amplitude of the light transmitted through the filter. These results are quite useful for analyzing various random and systematic effects. The general equations in this section will be evaluated in Secs. 3 and 4 for the specific case of combined random and systematic phase errors.

2.1 Diffraction from an array of pixels

The complex transmittance of an array of N pixels can be expressed as

$$T'(f) = \sum_{i=1}^{N} g_i(f-f_i)\,\exp[j\psi_i(f-f_i)] \tag{3}$$

where each pixel is described by an amplitude g_i and phase ψ_i function centered at position f_i in the filter plane. Eq. 1 also describes the complex amplitude of the light transmitted through an array of pixels. We will use the "prime" symbol to distinguish T'(f) the filter transmittance, from T(f) the complex amplitude of the transmitted light. The transmitted light diffracts to the complex amplitude in the correlation plane according to the Fourier transform relationship

$$c(x) = \sum_i A_i(x) = \mathcal{F}\{T(f)\} \tag{4}$$

where A_i is the complex amplitude of the diffraction pattern from the i'th pixel.

Diffraction from independently random pixels. The operations of expectation and summation are linear. Thus the expected value of the correlation amplitude is simply the sum of the expected values of the individual A_i. The observable of interest in correlators is typically the intensity $I_c(x)$. For the case in which each pixel transmittance, or equivalently, each A_i is mutually independent of the other the general expression for the expected correlation plane intensity is

$$\langle I_c(x)\rangle = \sum_i \sum_j \langle A_i(x) A_j^*(x)\rangle = |\langle c(x)\rangle|^2 + \sum_i \left[\, \langle |A_i|^2\rangle - |\langle A_i\rangle|^2 \,\right] \tag{5}$$

where $\langle\ \rangle$ represents the ensemble average or expectation operator. This expression shows that the expectation nearly separates into $|\langle c(x)\rangle|^2$ except for terms involving the autocorrelation of the filter plane pixels (i.e. terms for which i=j.) Under the same assumption of independent pixels, the most general expression for the squared far-field intensity is

$$\begin{aligned}
\langle I_c^2\rangle &= \sum_i \sum_j \sum_k \sum_l \langle A_i A_j^* A_k^* A_l\rangle \\
&= 2\left[\langle I_c\rangle^2 - |\langle c(x)\rangle|^4\right] + |\langle c(x)\rangle^2 + \sum_i \left(\langle A_i^2\rangle - \langle A_i\rangle^2\right)|^2 \\
&\quad + 4\,Re\left[\langle c^*(x)\rangle \sum_i \left(\langle |A_i|^2 A_i\rangle - \langle A_i^2\rangle\langle A_i^*\rangle + 2|\langle A_i\rangle|^2\langle A_i\rangle - 2\langle |A_i|^2\rangle\langle A_i\rangle\right)\right] \\
&\quad + \sum_i \Big[\langle |A_i|^4\rangle - 6|\langle A_i\rangle|^4 + 8\langle |A_i|^2\rangle|\langle A_i\rangle|^2 - |\langle A_i^2\rangle|^2 - 2\langle |A_i|^2\rangle^2 \\
&\qquad\qquad + 4\,Re\left(\langle A_i^2\rangle\langle A_i^*\rangle^2 - \langle |A_i|^2 A_i\rangle\langle A_i^*\rangle\right)\Big]
\end{aligned} \tag{6}$$

The derivation of these equations is given in ref. 8. The standard deviation of intensity $\sigma_I(x)$ is then directly found using the well known result

$$\sigma_I^2 = \langle I_c^2\rangle - \langle I_c\rangle^2 \tag{7}$$

2.2 Correspondence between correlation and expectation.

The correspondence follows from the fact that correlation involves integration and that integration is a smoothing or averaging operation. We show this for continuous signals in general.

In the classical correlator the signal s(x) is Fourier transformed to produce

$$S(f) = a'(f)\exp[j\phi(f)] = \mathcal{F}\{s(x)\} \tag{8}$$

where a'(f) is the positive valued amplitude spectrum and ϕ(f) is the phase spectrum. The transmittance chosen to recognize the signal is

$$T'(f) = [a'(f)]^n \exp[-j\phi(f)] \tag{9}$$

The exponent n allows one to consider any fractional power filter.[9] (Our previous analyses only considered phase-only filters, i.e. n = 1.) The correlation peak amplitude produced in response to the signal S(f) is

$$c(0) = \int_{-B_f/2}^{B_f/2} a(f)\, df = B_f\, \bar{a} \tag{10}$$

where a(f) = [a'(f)]$^{n+1}$ which is the magnitude of T(f). The overline in eq. 10 is used to indicate the spatial average (across the filter plane). This average is identical in form to the temporal average.[10,11] It indicates the importance of the signal spectrum on correlation. Correlation in the presence of systematic, modulation-dependent phase error can also be viewed as an average when expressed as

$$c(0) = \int_{-B_f/2}^{B_f/2} a(f)\exp\{j\,\delta\phi[\phi(f);k]\}\,df = B_f\,\overline{a\exp(j\,\delta\phi)} \tag{11}$$

where the phase error $\delta\phi$ has been expressed in the general form introduced in eq. 2. Eq. 11 can be rewritten as

$$c(0) = B_f \int_0^{\infty}\int_{-\pi}^{\pi} a\exp(j\delta\phi)\, p(a,\phi)\, d\phi\, da = B_f\,\langle a\exp(j\delta\phi)\rangle \tag{12}$$

As we noted in ref. 5, the spatial average in eq. 11 is equivalent to the integral in eq. 12. It takes the form of and approximates an ensemble average or expected value of the random variable contained between the brackets $< >$. The density function p(a,ϕ) that arises when performing a transformation of variables in integration represents a histogram of the ordered pairs (a,ϕ). As such, it approximates the joint probability density function of the random variable (a,ϕ). This correspondence between deterministic and random quantities is no different than intensity histogramming which is commonly used in image processing;[12] except that we are using complex-valued rather than real occurrences.

In order to appreciate how limited phase affects correlation we make the following simplifying approximations to eq. 11

$$c(0) \approx B_f\,\bar{a}\,\overline{\exp(j\delta\phi)} \approx \frac{B_f\,\bar{a}}{2\pi}\int_{-\pi}^{\pi}\exp(j\delta\phi)\,d\phi \tag{13}$$

A heuristic explanation for the first approximation is that phase more strongly influences correlation than amplitude. This approximation will generally be better if the values of phase error do not depend on the values of amplitude and is exactly true if a and ϕ are independent random variables. The second approximation is true if the phase ϕ is uniformly distributed over 2π. These approximations are especially useful in simplifying the analyses. The range of validity of these two approximations is considered futher in ref. 5. Ref. 5 also shows good correspondence between models using eq. 13 and computer simulated correlations using gray scale images.

3. MODELS OF THE CORRELATION PEAK AMPLITUDE

In this study we treat the transmittance (and likewise the spatial variation of the illumination) across the aperture of each pixel as a constant. This is reasonable as long as the spatial variation of the illumination is at most half the sampling rate of the pixel array. Under these assumptions the transmittance in eq. 9 can be simplified to

$$T'(f) = \sum_i \left[a_i'\right]^n \exp(-j\phi_i)\ \mathrm{rect}\left(\frac{f-f_i}{w}\right) \tag{14}$$

where a_i and ϕ_i are constants across the pixel aperture of width w. From the Fourier transform of T(f) the peak amplitude of the correlation peak is found to be

$$c(0) = \sum_i a_i \exp(j\delta\phi_i) \approx Nw\bar{a}\,\overline{\exp(j\delta\phi)} \tag{15}$$

where the approximations from eq. 13 have been used.

Statistical expressions. For the analysis of phase errors we will consider the amplitudes to be non-random and the phases to be composed of a non-random component plus a random component of zero mean (i.e. at each pixel the phase is of the form $\phi = \phi_s + \phi_r$ where the subscripts indicate systematic and random phase.) The random components of phase are assumed to be independent and identically distributed random variables for this analysis. As such, the expected value of the random phasor components at each pixel will be identical. Under these conditions the expected value of the peak correlation amplitude becomes

$$\langle c(0)\rangle = M(1)\sum_i a_i \exp(j\phi_{is}) = Nw\bar{a}\,\overline{\exp(j\phi_s)}\,M(1) \approx Nw\bar{a}M_s(1)M(1) \tag{16}$$

where

$$M(\omega) = \langle \exp(j\omega\phi)\rangle = \int_{-\infty}^{\infty} \exp(j\omega\phi)\,p_\phi(\phi)\,d\phi \tag{17}$$

is referred to as the characteristic function[11] of the random variable ϕ and where $p_\phi(\phi)$ is the probability density function of ϕ. The approximation in eq. 16 again follows from eq. 13. Eqs. 12 and 13 together are also used to approximate the spatial average of the phasors as the characteristic function $M_s(1)$.

The expected intensity of the correlation peak follows from eqs. 5 and 16 as

$$\langle I(0)\rangle = |\langle c(0)\rangle|^2 + Nw^2q\,\overline{a^2} = (Nw\bar{a})^2\left(pp_s + qZ/N\right) \tag{18}$$

where the shorthand

$$\begin{aligned} p &= M^2(1) \\ p_s &= M_s^2(1) \\ q &= 1 - p \\ Z &= \frac{\overline{a^2}}{\bar{a}^2} \end{aligned} \tag{19}$$

is used. The standard deviation of the correlation peak is found by evaluating eq. 7 and using eqs. 16 and 18 to get

$$\sigma_I^2(0) = (Nw\bar{a})^4\left\{2\frac{Z}{N}\left[q+(d-p)d_s\right]pp_s + \left[\frac{Z}{N}(d-p)d_s\right]^2 + 4\frac{Z_3}{N^2}\left[2p-d-1\right]pp_s + \frac{Z_4}{N^3}\left[4p-6p^2+4pd-d^2-1\right]\right\} \tag{20}$$

where we use the additional shorthand notation

$$\begin{aligned} d &= M(2) \\ d_s &= M_s(2) \\ Z_l &= \frac{\overline{a^l}}{\bar{a}^l} \end{aligned} \tag{21}$$

4. COMPARISON OF THE EFFECT OF VARIOUS PHASE ERRORS ON CORRELATION AMPLITUDE

In eq. 16 the effect of each type of phase error can be seen to produce a loss in amplitude that is proportional to the characteristic function of that phase error. Their combined effect is multiplicative. For purposes of comparison it is useful if each characteristic function can be written in terms of a single argument that describes the amount of phase match or mismatch. We have done this is in Table I where we have chosen to use k, the maximum phase match as the single parameter. The third column shows the same results in terms of the parameters that are typically used. For example, uniformly distributed

random noise is usually described in terms of its total spread ν. This spread describes the amount of phase mismatch rather than phase match and is in units of radians. The gaussian has been rewritten in terms of k by relating ν for the uniform distribution to its standard deviation. This proportionality has then been used to substitute out σ in the characteristic function for the gaussian. The fourth column shows that many types of phase errors affect correlation amplitude to a similar (if not exact) degree when expressed in terms of k. The saturated characteristic is especially interesting in that its effect can be described as a linear combination of k multiplied by c(0;1) the correlation amplitude without phase errors and 1-k multiplied by $c_n(0;k)$ the correlation amplitude for the non-unity slope characteristic.

Phase Characteristic	Phase Error	$\frac{\lvert c(0)\rvert}{Nw\bar{a}}$	$\frac{\lvert c(0;k)\rvert}{Nw\bar{a}}$
SYSTEMATIC	$\delta\phi$		
Saturated for $\frac{\lvert\phi\rvert}{\pi} \geq k$	$0 \quad ; \lvert\phi\rvert \leq k\pi$ $\phi - \frac{k\pi\phi}{\lvert\phi\rvert} \quad ; \lvert\phi\rvert \geq k\pi$		$k + (1-k)\mathrm{sinc}(1-k)$
Non-unity slope of k	$(1-k)\phi$		$\mathrm{sinc}(1-k)$
Quantized in m levels	$\mathrm{mod}\left(\phi + \frac{\pi}{m}, \frac{2\pi}{m}\right)$	$\mathrm{sinc}\left(\frac{1}{m}\right)$	$\mathrm{sinc}(1-k)$
RANDOM	$p_\Phi(\phi)$		
Uniform of spread ν	$\frac{1}{\nu}$ for $\frac{-\nu}{2} \leq \phi \leq \frac{\nu}{2}$	$\mathrm{sinc}\left(\frac{\nu}{2\pi}\right)$	$\mathrm{sinc}(1-k)$
Gaussian of standard deviation σ	$\frac{1}{\sqrt{2\sigma^2}} \exp\left(\frac{-\phi^2}{2\sigma^2}\right)$	$\exp\left(-\frac{\sigma^2}{2}\right)$	$\exp\left[-\frac{\pi^2}{6}(1-k)^2\right]$

Table 1. Definitions of various phase errors and relation of their effects through the parameter k.

5. MODEL OF PEAK-TO-NOISE RATIO

In order to illustrate the combined effects of random and systematic phase errors more clearly we develop a model expression for the peak correlation amplitude to noise ratio (PNR).[9,13] We start with the general expression of

$$PNR = \sqrt{\frac{\langle I_c(0)\rangle}{\frac{1}{B_x - \Delta_x}\left[\int_{-B_x/2}^{B_x/2} \langle I_c(x)\rangle\, dx - \int_{-\Delta_x/2}^{\Delta_x/2} \langle I_c(x)\rangle\, dx\right]}} \tag{22}$$

where the denominator represents the root-mean-square amplitude across a correlation plane of spatial bandwidth B_x which excludes a small region of width Δ_x centered around the correlation peak.

We have chosen to model the filter plane SLM as an array of equally spaced pixels of pitch Δ_f. This leads to a non-redundant bandwidth $B_x = \Delta_f^{-1}$; i.e.the spacing between diffraction orders of the array. The width $\Delta_x = B_x / N$ corresponds to the resolution of an N element array and these definitions also lead to $B_f = N\Delta_f = \Delta_f^{-1}$. These choices have been made specifically so that our model closely correspond to typical fast Fourier transfom based computer simulations that represent each filter plane pixel with one sample.[14] Consistent with these assumptions we approximate the second integral in the denominator of eq. 22 as $\Delta_x \langle I_c(0)\rangle$. The first integral is the energy in the central diffraction order. It can be related to the energy in the filter plane through Parseval's theorem. The total energy in the filter plane and the correlation plane is

$$E_f = \int_{-B_f/2}^{B_f/2} a^2(f)\, df = B_f \overline{a^2} = E_c \tag{23}$$

The energy in the central diffraction order is found by recognizing that the ratio of energy in each diffraction order is determined from the Fourier series of a square wave grating of duty cycle D = w/Δ_f where w is the width of each pixel. Therefore the first integral in eq. 22 can be written as D^2E_f. Using these results eq. 22 can be written as

$$PNR = \sqrt{\frac{(N-1)\langle I_c(0)\rangle}{N^2 w^2 \overline{a^2} - \langle I_c(0)\rangle \Delta_f}} \tag{24}$$

Using eq. 18 for $\langle I_c \rangle$ in eq. 24 gives

$$PNR = \sqrt{\frac{(N-1)(pp_s + qZ/N)}{Z - (pp_s + qZ/N)}} \tag{25}$$

Evaluation of the model. Fig. 1a shows the effect on PNR of combined non-unity slope (described by parameter k) and uniformly distributed random phase (described by the parameter k_r.) It has been calculated for a 128x128 SLM, i.e. N=16,384 and a transmitted amplitude spectrum of Z=6.23. These parameters correspond to those for a companion simulation on the effect of systematic phase errors.[5] The curve for k_r=1 is identical to our modeled results from that study. The term pp_s is typically much larger than the term qZ/N. This can also be seen in fig. 1a. which plots (as dotted lines) eq. 25 with the term qZ/N set to zero. The dotted lines are barely distinguishable except for where the PNR is less than 3. Fig. 1b quantifies these differences as a relative error with respect to PNR as given by eq. 25.

For the example plotted in fig. 1 the denominator has a small effect. This is because the term subtracted from Z is never greater than unity. The minimum value of Z is unity which occurs if the amplitude of the optical field that is transmitted through the filter is constant. This situation occurs if an inverse filter, i.e. a fractional power filter of n = -1, is used. In this instance PNR will be infinite when there is no phase error. This points out one idealization in our model. It is that the signal amplitude and the filter amplitude (for n=1) are identical. This will not be the case for a signal in noise or clutter, or for scale or rotation errors. Evidence (via computer simulation) that these models are applicable for images in clutter by adjusting the value of Z is given in ref. 5.

PNR depends on the expected value of $I_c(0)$ the peak intensity. One way to estimate the sensitivity of PNR to phase error induced intensity fluctuations is to perturb the expected intensity in eq. 24 by $\pm \sigma_I(0)$ from eq. 20 to get

$$PNR \pm \delta PNR = \sqrt{\frac{(N-1)\left[pp_s + qZ/N + \sigma_I(0)/(Nw\bar{a})^2\right]}{Z - \left[pp_s + qZ/N + \sigma_I(0)/(Nw\bar{a})^2\right]}} \tag{26}$$

Fig. 2 shows $\pm$ δPNR/PNR the relative fluctuation for the PNR curves in Fig. 1. The smallest fluctuations are found for values of k and k_r near unity. As the values of these parameters decrease, the fluctuation increases, but only to a point. Below that point the fluctuation decreases. The reason for the decrease is that the intensity of the correlation peak is being dominated by noise or speckle background. The turning point occurs when pp_s (which is proportional to $|c(o)|^2$ is of the same order of magnitude as qZ/N (which is proportional to the average intensity of the speckle background.) The light lines in Fig. 2 are approximations in which all but the first of the four terms in braces in eq. 20 have been omitted in the calculations. The differences are barely distinguishable. This differences are primarily due to the second of the four terms. We originally chose to use values of $Z_3=Z^{1.5}$ and $Z_4=Z^2$ but numbers a few orders of magnitude larger caused no noticeable difference. What is interesting is that correlation performance can be predicted with such a small amount of specific information about the amplitude. In our model, the normalized second order moment of amplitude $Z=Z_2$ (see eqs. 19 and 21) is required but almost no useful information is provided by the third and fourth moments Z_3 and Z_4. It is also interesting that the fluctuations in PNR are quite low even when the PNR has been reduced by random phase by over a factor of ten. This is of course due to the large number of pixels available in current SLMs.

Fig. 3a shows PNR for specific ratios of k_r divided by k. Fig. 3b shows the corresponding fluctuations. The combination of errors is noticeably, though not catastrophically reduced by adding random phase error that is equivalent to the systematic error. A k of .5 also corresponds to the systematic phase errors produced by a binarized phase filter phase (see Table 1.) For k_r/k = 1 the spread v of the random phase errors equals to the quantization. The PNR is only reduced about 40 % at k = .5 by this amount of random phase. The positive fluctuation is around 3 % and can be read from either fig. 2 or fig. 3b. The curves

can be used to consider any amount of quantization by using the mapping k = (m-1)/m. Note also in fig. 3b that the fluctuation is essentially zero for $k_r/k = 2$ around k = .5 . At this point the fluctuation is again dominated by speckle.

6. CONCLUSIONS

We have developed a model that describes the performance of optical correlators subject to a combination of phase errors. The effect of various types of phase errors were related through a single parameter k. Other systematic and random phase errors of interest can be described by different characteristic functions and used in the performance model. It has been possible to develop such a simple model by decoupling any dependence between the phases and the amplitudes in the filter plane. When this is done the effects of amplitude is totally contained in the single parameter Z which the normalized second moment of the amplitude spectrum. This approximate model provides much insight and is useful at the early stages of design. It should be considered prior to performing exhaustive simulations and design studies on the effects of phase errors. Part of this analysis should consider the values of Z for the objects in the image training.

Of course, for many SLMs the values of phase and amplitude are coupled.[15] We did not consider this problem here. However, it seems reasonable that there are alternate evaluations or approximations of eq. 12 that can lead to a model of nearly comparable usefulness and simplicity.

7. ACKNOWLEDGEMENTS

RWC acknowledges the sponsorship of the Advanced Research Projects Agency (ARPA) through U.S. Air Force Rome Laboratory contract F19628-92-K-0021.

8. REFERENCES

1. Laser Speckle and Related Phenomena, second edition, J.C. Dainty, ed., Springer, Berlin (1984).
2. J.L. Horner and P.D. Gianino, "Effects of quadratic phase distortion on correlator performance," *Appl. Opt.*, 31, 20, 3876-3878. (10 July 1992)
3. R.W. Cohn and J.L. Horner, "Limited phase modulation and its effect on phase-only correlation," *Proc. SPIE* 2237-16. (4 April 1994)
4. W.J. Smith, *Modern Optical Engineering*, Ch. 11. McGraw-Hill, New York. (1966)
5. R.W. Cohn and J.L. Horner, "Effects of systematic phase errors on phase-only correlation," *Appl. Opt.* (to appear 1994)
6. R.W. Cohn and R.J. Nonnenkamp, "Statistical moments of the transmittance of phase-only spatial light modulators," *Proc. SPIE* 1751, 289-297. (1992)
7. R.W. Cohn, "Random phase errors and pseudorandom phase modulation of deformable mirror spatial light modulators," *Proc. SPIE* 1772, 360-368. (1992)
8. R.W. Cohn and M. Liang, "Approximating fully complex spatial modulation with pseudorandom phase-only modulation," *Appl. Opt.* (to appear 1994)
9. B.V.K.V. Kumar and L.G. Hassebrook, "Performance measures for correlation filters," *Appl. Opt.*, 29, 20, 2997-3006. (10 July 1990)
10. A. Yariv, *Optical Electronics*, 4th ed. Chs. 1 and 10. Saunders College Publishing, Philadelphia. (1991)
11. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 1st ed. McGraw-Hill, New York. (1965)
12. R.C. Gonzalez and P. Wintz, *Digital Image Processing*, Addison-Wesley, Reading, MA (1977).
13. J.L. Horner, "Metrics for assessing pattern-recognition performance," Appl. Opt., 31, 2, 165-166. (10 January 1992)
14. A.V. Oppenheim and R.W. Schafer, Digital Signal Processing, Sec. 5.5, Prentice-Hall, Englewood Cliffs (1975).
15. R.D. Juday, "Correlation with a spatial light modulator having phase and amplitude cross coupling," *Appl. Opt.*, 28, 22, 4865-4869. (15 November 1989)

Errata on fig. 3 and related text: Replace k_r/k the ratio of phase match with $(1-k_r)/(1-k)$ the ratio of phase mismatch.

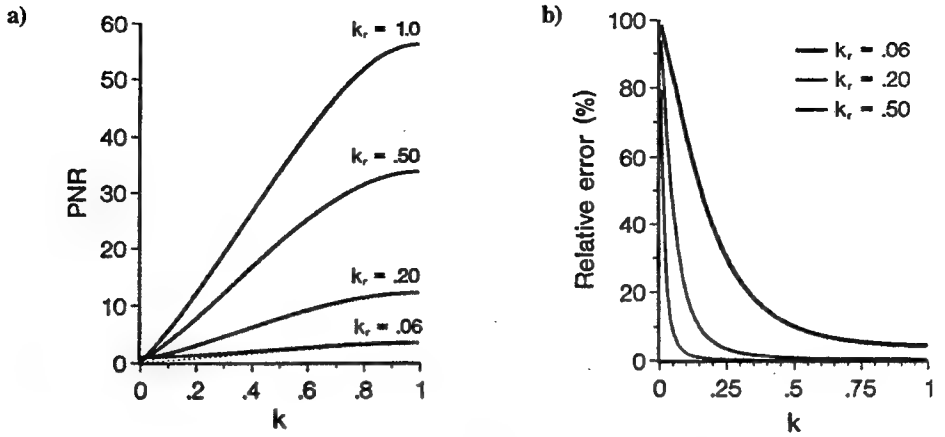


Fig. 1. PNR as a function of systematic and random k. a) PNR heavy lines, approximated PNR dotted lines. b) Relative error between PNR and approximated PNR.

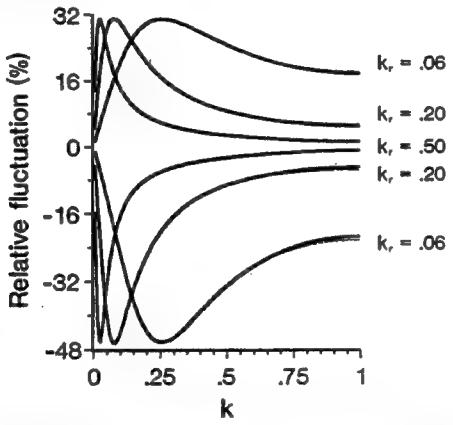


Fig. 2. Fluctuation in PNR due to perturbation of intensity by plus and minus the standard deviation of intensity. Fluctuation shown as heavy lines, approximated fluctuation shown as light lines.

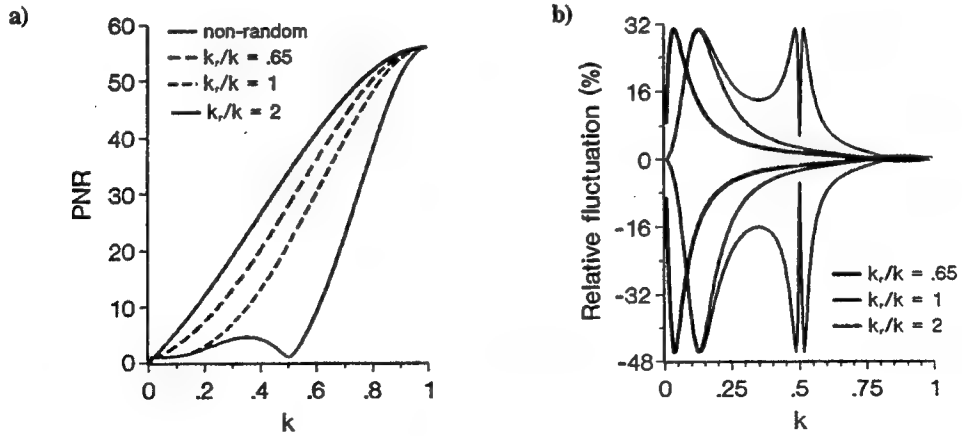


Fig. 3. a) PNR shown as a function of k and k_r/k. b) Relative fluctuation of PNR due to perturbation of intensity by plus and minus the standard deviation of intensity.

Performance models of correlators with random and systematic phase errors

Robert W. Cohn, MEMBER SPIE
University of Louisville
Department of Electrical Engineering
Louisville, Kentucky 40292
E-mail: rwcohn01@ulkyvm.louisville.edu

Abstract. Hybrid correlators are composed of numerous nonideal electronic and optical components that, to one degree or another, limit performance through unintended transformations of signals. Many of these effects show up as phase errors at a spatial light modulator (SLM) plane. The errors can be described as random variables, or as systematic offsets from the correct phases, as appropriate. Sources of systematic phase errors include quantizing circuits, incorrect or nonlinear amplifier gain, limited range phase modulators and residual phase modulation of amplitude-mostly SLMs. Random phase errors arise from electronic noise and fabrication variations of SLMs. Several systematic and random filter plane errors are related through a single parameter that describes the amount of phase mismatch. A model of peak-to-noise ratio (PNR) is also presented that describes the combined effects of random and systematic errors. This expression contains the products of two functions, one that depends only on systematic, the other on random, phase mismatch. PNR is also a function of the number of pixels in the filter plane modulator and a normalized moment of the amplitude of the image spectrum. The model is useful for developing phase error budgets for correlation systems.

Subject terms: optical correlation; spatial light modulators; phase errors; correlation metrics; pattern recognition; diffractive optics; laser speckle.

Optical Engineering 34(6), 1673-1679 (June 1995).

1 Introduction

Optical correlation is highly dependent on the amount of phase mismatch or wavefront error across the filter plane. If the filter perfectly conjugates or matches the incident wavefront, then essentially all the energy is focused into a single bright spot at the correlation plane. If the filter poorly matches the incident wave, then the filter behaves like a diffuser in that the correlation intensity is greatly reduced and that a broad background of scattered light is produced akin to a speckle pattern.[1] It has been noted for various types of phase errors that small to moderate errors, say less than a quarter wave, only slightly reduce correlation intensity and only slightly increase the level of background noise in the correlation plane.[2,3] A quite similar situation is that the resolution of imaging systems is nearly diffraction limited if the total aberrations are less than one quarter wavelength.[4] Just as in imaging, the critical issue in optical correlation is the total amount of phase error rather than the functional description of the phase error. The physical reason for this is that the formation of a correlation peak is caused by the coherent superposition of a large number of wavefronts. Each cell or pixel in the filter plane produces at the correlation plane a wavefront with a specific amount of phase error and thus, the amount of filter plane phase error, rather than its spatial distribution controls the intensity of the correlation peak.

We have recently begun modeling the coherent formation of the correlation peak as a statistical process. The model is applicable even when the phase errors are nonrandom (i.e., systematic). Our earlier work on phase errors considered systematic[5] and random[6–8] phase errors individually. In this paper, we now are able to make direct comparisons between the relative effects of various systematic and random phase errors. The model can also be used to calculate the combined effects of systematic and random errors. Whereas the earlier models focused on phase-only filters, the models here apply to fully complex fractional power filters[9] (including the phase-only filter).

2 Correlation in the Presence of Filter Phase Errors

To focus on the effect of filter plane phase errors on correlation we model the idealized situation of a signal being correlated with a distorted version of itself. All distortions are produced by the spatial light modulator at the filter plane of the classical correlator and there are no other sources of noise or clutter. The optical correlator is translation invariant so it is sufficient to consider only the case where there is no coordinate shift between the signal and the impulse response of the filter. Thus, in this analysis the correlation plane response $c(x)$ peaks at $x=0$.

Paper 20094 received Sep. 24, 1994; revised manuscript received Jan. 12, 1995; accepted for publication Jan. 16, 1995. This paper is a revision of a paper presented at the SPIE conference on Advances in Optical Information Processing VI, April 1994, Orlando, FL. The paper presented there appears (unrefereed) in SPIE Proceedings Vol. 2240.

The Fourier plane spectrum of the input plane signal $s(x)$ is written

$$S(f)=a'(f)\ \exp[j\phi(f)]\ , \tag{1}$$

where $a'(f)$ is the positive valued amplitude spectrum and $\phi(f)$ is the phase spectrum. The filter transmittance is of the form

$$T'(f)=[a'(f)]^n\ \exp[-j\phi(f)+j\delta\phi(f)]\ , \tag{2}$$

where $\delta\phi$ is the phase error introduced by the SLM and the exponent n allows one to consider any fractional power filter,[9] including the phase-only filter for $n=0$. It is worth emphasizing, that although the amplitude and phase of many current SLMs are known to be coupled,[10] we are not considering this possibility in this analysis. Multiplying the two spectra [Eq. (1) and (2)] together produces the transmitted spectrum

$$T(f)=[a'(f)]^{n+1}\ \exp[j\delta\phi(f)]\equiv a\ \exp(j\delta\phi)\ , \tag{3}$$

where $a=|T|$.

The results developed in the following can be directly applied to a 2-D array of N pixels for which the pixels are regularly spaced with pitch Δ_{fx},Δ_{fy} in the frequency plane coordinates f_x and f_y, and for which each pixel has clear area w. To simplify explanation, however, the SLM is modeled as a 1-D array of N pixels, each of finite aperture of width w and regularly spaced with pitch Δ_f. (For examples of applying 1-D models to 2-D problem see Refs. 5 and 8.) If the illumination $S(f)$ contains no spatial frequencies that exceed half the sampling rate of the pixel array it is reasonable to assume that the transmitted spectrum can be described as a sampled function of the form

$$T(f)=\sum_{i=1}^{N} a_i\ \exp(j\delta\phi_i)\ \text{rect}\left(\frac{f-\Delta-i\Delta_f}{w}\right)\ , \tag{4}$$

where a_i and $\delta\phi_i$ are the transmitted amplitude and phase at the i'th pixel. Value of the offset Δ is set to center the transmitted spectrum on the optical axis. The transmitted light diffracts into the complex amplitude distribution in the correlation plane according to the Fourier transform relationship

$$c(x)=\mathcal{F}[T(f)]\equiv\sum_i A_i(x)\ , \tag{5}$$

where A_i is the complex amplitude of the diffraction pattern from the i'th pixel. The amplitude of the resulting correlation peak is

$$c(0)=\sum_i a_i\ \exp(j\delta\phi_i)=Nw\overline{a\ \exp(j\delta\phi)}\ . \tag{6}$$

Centering of the transmitted spectrum (via Δ) in Eq. (4) eliminates a phase shift in Eq. (6) that is unimportant in our analysis. The second equality in Eq. (6) shows that the peak amplitude corresponds to the spatial average (indicated by overline) of $T(f)$ multiplied by Nw the active transmitting area of the SLM. If there are no phase errors the correlation peak is a direct measure of $\bar{a}$ the average amplitude of $T(f)$. Equation 6 makes clear that the peak amplitude is independent of the spatial distribution of the phase errors across the filter plane, as initially discussed in Sec. 1.

2.1 *Influence of Systematic Phase Errors on the Correlation Peak*

One class of systematic phase errors is modulation-dependent errors.[5] These phase errors can be described as an explicit, nonrandom function of the form $\delta\phi(\phi;k)$, where phase error is a function of the value of the signal phase ϕ and a parameter k that selects one function from a class of similar functions. For example, if the SLM only produces $-k\phi$ for each value ϕ from the signal spectrum then the phase error is

$$\delta\phi(\phi;k)=(1-k)\phi\ . \tag{7}$$

We refer to this type of phase error as a gain or slope error. The parameter k indicates to what degree the SLM phase matches the phase of the signal spectrum and 1-k indicates the magnitude of the phase errors. Various phase errors that result from systematic mappings of the desired phase ϕ to the actual phase produced by the SLM are defined in Table 1. The first column of the table describes the mapping relationship and the second column gives the expression for systematic phase errors (including nonunity slope error) as functions of k. Note for the case of phase quantized SLMs, which are commonly described in terms of the number of levels of quantization m, we have defined an equivalent phase match parameter k using the relationship[5] $m=1/(1-k)$.

A reasonable first-order approximation to Eq. (6), the correlation peak amplitude is[5]

$$c(0;k)\approx\frac{Nw\bar{a}}{2\pi}\int_{-\pi}^{\pi}\exp[j\delta\phi(\phi;k)]\ d\phi\ , \tag{8}$$

which follows from modeling the amplitudes a_i and the phases ϕ_i of the transmitted spectrum as random variables. Specifically, the approximation follows if the amplitudes and phases are statistically independent of each other, and if the phases ϕ_i are uniformly distributed over 2π. The errors introduced by this approximation can be surprisingly small as compared to what one actually finds when digital simulating the optical correlation process using image of real-world objects, such as tanks and faces.[5] The closeness of the approximation to the exact result depends on the actual statistics or histogram of the occurrences of (a_i,ϕ_i). Equation (8), however, is often preferable to an exact result because it can be evaluated in closed form for many cases of interest. Closed-form expressions for Eq. (8), the effect of four specific types of systematic phase error on $c(0)$ are given in the third and fourth columns of Table 1. The third column is expressed in terms of commonly used parameters, e.g., m, if available. The fourth column is written in terms of k. Expressing correlation peak magnitude in terms of a single parameter is used to more directly compare the effect of each systematic (and also random) phase error in Sec. 2.3.

Note also that the magnitude of Eq. (8) is unchanged if there is a constant offset in the phase error. This has been used to simplify some of the expressions for phase error in Table 1. For the case of a phase quantized SLM, the phase error varies between 0 and $2\pi/m$. One might initially assume that quantizing so that the phase error varies between $-\pi/m$ and π/m would produce a larger correlation peak. Because the constant $\exp(j\pi/m)$ can be factored outside the

Table 1 Definitions of various systematic phase errors and relation of their effects through the parameter k.

PHASE MAPPING	$\delta\phi(k)$	$\lvert c(0)\rvert / Nw\bar{a}$	$\lvert c(0;k)\rvert / Nw\bar{a}$
Saturated at $\pm k\pi$	$0 \quad ; 0 \le \lvert\phi\rvert \le k\pi$ $\phi - k\pi\,\mathrm{sgn}\,\phi \quad ; k\pi \le \lvert\phi\rvert \le \pi$		$k + (1-k)\,\mathrm{sinc}\,(1-k)$
Non-unity slope of k	$(1-k)\phi$		$\mathrm{sinc}\,(1-k)$
Quantized in m levels	$\mathrm{mod}\,[\phi, 2\pi(1-k)]$	$\mathrm{sinc}\,\frac{1}{m}$	$\mathrm{sinc}\,(1-k)$
Binarized at 0 and $k\pi$	$\phi \quad ; 0 \le \phi \le k\pi$ $(1-k)\phi \quad ; k\pi \le \phi \le 2\pi$	$\mathrm{sinc}\,\frac{1}{2}\,\sin\frac{k\pi}{2}$	$k\,\mathrm{sinc}\,\frac{k}{2}$

Table 2 Definitions of various random phase errors and relation of their effects through the parameter k.

RANDOM ERRORS	pdf $(\delta\phi)$	$\lvert\langle c(0)\rangle\rvert / Nw\bar{a}$	$\lvert\langle c(0;k)\rangle\rvert / Nw\bar{a}$
Uniform of spread v	$\frac{1}{v}\,\mathrm{rect}\,\frac{\delta\phi}{v}$	$\mathrm{sinc}\,\frac{v}{2\pi}$	$\mathrm{sinc}\,(1-k)$
Gaussian of standard deviation σ	$\frac{1}{\sqrt{2}\sigma}\,\exp\frac{-1}{2}\left(\frac{\delta\phi}{\sigma}\right)^2$	$\exp\frac{-\sigma^2}{2}$	$\exp\frac{-\pi^2}{6}(1-k)^2$

integral in Eq. (8), however, there is no change in the magnitude of the correlation peak. This property has been used to remove two additive terms of π/m from the expression for phase error.

2.2 *General Analysis of the Influence of Random Phase Errors on the Correlation Peak*

Another class of phase errors is random. We specifically consider the case in which the phase errors are independent and identically distributed random variables, and the amplitudes are nonrandom. Furthermore, spatial variation of the transmitted spectrum over the aperture of a pixel is neglected [see Eq. (4)]. We derive expressions for $\langle c(0)\rangle$, the expected peak correlation amplitude; $\langle I_c(0)\rangle$, the expected intensity of the correlation peak; and $\sigma_I(0)$, its standard deviation, where $\langle\ \rangle$ represents the ensemble average or expectation operator.

The desired expressions follow from equations developed by Cohn and Liang that describe diffraction from an array of pixels in an illumination plane to a Fourier transform plane.[8] The equations were developed under the assumption that the light transmitted through any pixel is statistically independent of the light at any other pixel. No other assumptions were made about the spatial or statistical properties of the light transmitted through the SLM. The equations are more general than the present analysis in that (1) the SLM transmittance and the transmitted spectrum amplitude can vary with position across the aperture of a pixel, (2) both amplitude and phase can be random variables, and (3) the statistics from pixel to pixel can be nonidentical.

These general equations are rewritten here to specifically represent diffraction from the filter plane to the correlation plane. Equation (5) is the general form of the correlation plane amplitude. [The additional spatial description provided by Eq. (4) is not needed at this point. It will be used in Sec. 2.3.] Taking the expectation of Eq. (5) gives the expected correlation amplitude

$$\langle c(x)\rangle = \sum_i \langle A_i(x)\rangle \ . \tag{9}$$

The general expression for the expected correlation plane intensity is

$$\langle I_c(x)\rangle = \sum_i \sum_j \langle A_i(x) A_j^*(x)\rangle$$

$$= |\langle c(x)\rangle|^2 + \sum_i [\langle |A_i|^2\rangle - |\langle A_i\rangle|^2] \ . \tag{10}$$

This expression separates into $|\langle c(x)\rangle|^2$ plus additional terms that result from the term $A_i(x)$ not being statistically independent from itself (i.e., terms in the double summation for which $i=j$). The general expression for the standard deviation of the correlation plane intensity distribution under the assumption that the $A_i(x)$ are statistically independent is

$$\sigma_I^2 = \langle I_c\rangle^2 - 2|\langle c\rangle|^4 + |\langle c\rangle^2 + \sum_i (\langle A_i^2\rangle - \langle A_i\rangle^2)|^2$$

$$+ 4\mathrm{Re}[\langle c^*\rangle \sum_i (\langle |A_i|^2 A_i\rangle - \langle A_i^2\rangle\langle A_i^*\rangle + 2|\langle A_i\rangle|^2\langle A_i\rangle - 2\langle |A_i|^2\rangle\langle A_i\rangle)]$$

$$+ \sum_i [\langle |A_i|^4\rangle - 6|\langle A_i\rangle|^4 + 8\langle |A_i|^2\rangle |\langle A_i\rangle|^2 - |\langle A_i^2\rangle|^2 - 2\langle |A_i|^2\rangle^2$$

$$+ 4\ \mathrm{Re}(\langle A_i^2\rangle\langle A_i^*\rangle^2 - \langle |A_i|^2 A_i\rangle\langle A_i^*\rangle)] \ , \tag{11}$$

which was originally derived by evaluating

$$\langle I_c^2\rangle = \sum_i \sum_j \sum_k \sum_l \langle A_i A_j^* A_k^* A_l\rangle = \langle I_c\rangle^2 + \sigma_I^2(x) \ . \tag{12}$$

The specific results for the on-axis amplitude, intensity, and standard deviation when $T(f)$ is of the form of Eq. (4) are not given until Sec. 2.3, where they are directly found by setting the systematic errors to zero in subsequent equations that describe combined random and systematic errors.

2.3 *Influence of Random and Systematic Phase Errors on the Correlation Peak*

The deviation of phase of the transmitted wavefront at the filter plane is of the form $\delta\phi = \delta\phi_s + \delta\phi_r$, where the subscripts indicate the systematic and random phase errors. Using the preceding results and assumptions the expected peak correlation amplitude becomes

$$\langle c(0)\rangle = w\langle \exp(j\delta\phi_r)\rangle \sum_i a_i \exp(j\delta\phi_{is})$$

$$\approx Nw\bar{a}\ \overline{\exp(j\delta\phi_s)}\ \langle \exp(j\delta\phi_r)\rangle \ . \tag{13}$$

The approximation in Eq. (13) follows from Eqs. (6) and (9). Equation (13) shows that the correlation amplitude is proportional to the product of the average systematic and the average random error phasors.

2.3.1 *Influence of random phase errors alone*

If the systematic phase errors are set to zero, then Eq. (13) describes the effect of specific random phase errors on correlation peak amplitude. Results for two specific types of random phase errors, uniform and Gaussian distributed, are given in Table 2. These distributions are commonly described in terms of the spread v for the uniform and the standard deviation σ for the Gaussian. The probability density function $\text{pdf}(\delta\phi)$ is given in the second column of Table 2 and $\langle c(0)\rangle$ is given in the third column in terms of v and σ. Each parameter has been transformed into k, the degree of phase match (in units of wavelengths), and these results are presented in the fourth column. For the Gaussian case, σ, the standard deviation for the Gaussian distribution, has been defined in terms of k by using the proportionality between spread v and σ_u, standard deviation of the uniform distribution ($\sigma_u^2 = v^2/12$).

2.3.2 *Comparing effects of various phase errors in terms of the phase match parameter k.*

Both systematic and random average error phasors can be evaluated in closed form for a number of specific cases of interest. Six cases are shown in Tables 1 and 2. Comparisons can easily be made, especially for the cases of nonunity slope, quantization, and uniform random phase errors, which have results of identical form. For the saturated case, the effect can be described as a linear combination of k multiplied by $c(0;1)$, the correlation amplitude without phase errors, and 1-k multiplied by $c_n(0;k)$, the correlation amplitude for the nonunity slope characteristic. An additional correspondence between the quantized and binarized cases is brought out in the third column of Table 1. The result for quantized phase was calculated assuming that quantization is in steps of $2\pi/m$, whereas the result for binarized phase considers the phase levels to be other than 0 and π. The third column for the binarized case is written in the form of the correlation amplitude for quantized case when $m = 2$ multiplied by a sine function. Written in this way, it can be seen that the reduction in correlation amplitude is slight for a separation between phase levels that is somewhat less than π. In all cases, as k decreases from unity, correlation amplitude decreases from its maximum of $Nw\bar{a}$.

2.3.3 *Expected peak intensity and its standard deviation*

The expected intensity of the correlation peak follows from Eqs. (10) and (13) as

$$\langle I_c(0)\rangle = |\langle c(0)\rangle|^2 + Nw^2 q\overline{a^2}$$

$$= (Nw\bar{a})^2(pp_s + qZ/N) \ , \tag{14}$$

where the shorthand

$$p = \langle \exp(j\delta\phi_r)\rangle^2,$$

$$p_s = \overline{\exp(j\delta\phi_s)}^{\,2},$$

$$q = 1 - p,$$

$$Z = \frac{\overline{a^2}}{\bar{a}^2} \ , \tag{15}$$

is used. In Eq. (14), the only terms that depend on phase error are p_s, p, and q. Note that Tables 1 and 2 give the dependence of p_s and p on the phase match parameter k because these terms are simply squares of the expressions in the fourth columns of the tables. For purposes of analyzing the effects of combined phase errors we will distinguish between the parameters describing systematic and random phase errors using the symbols k and k_r, respectively.

Equation (14) is composed of two additive terms. The first term is the intensity of the desired coherent correlation peak in Eq. (13). The second term can be identified with the average intensity of incoherent noise or speckle background.[1] The relative magnitude of the coherent term will exceed that of the incoherent term as long as the phase errors and Z, a normalized moment of the transmitted amplitude spectrum, are not too large. The values of the parameter Z range from a minimum value of unity if the amplitude spectrum is uniform across the SLM to a maximum value of N if only one of the N pixels is transmitting. A unity value of Z corresponds to the case of inverse filtering; i.e., a fractional power filter with $n = -1$.

The standard deviation of the correlation peak is found by evaluating Eq. (11) along with the results in Eqs. (13) to (15) to get

$$\sigma_I^2(0) = (Nw\bar{a})^4 \left\{ 2\frac{Z}{N}[q + (d-p)d_s]pp_s + \left[\frac{Z}{N}(d-p)d_s\right]^2 \right.$$

$$\left. + 4\frac{Z_3}{N^2}(p - d - q)pp_s + \frac{Z_4}{N^3}(3p - 6p^2 + 4pd - d^2 - q) \right\}, \tag{16}$$

where we use the additional shorthand notation

$$d = \langle \exp(j2\delta\phi_r)\rangle,$$

$$d_s = \overline{\exp(j2\delta\phi_s)},$$

$$Z_I=\frac{\overline{a^I}}{\overline{a^I}} . \tag{17}$$

3 Model of Peak-to-Noise Ratio

To illustrate the combined effects of random and systematic phase errors more clearly we develop a model expression for the peak correlation amplitude to noise ratio[9,11] (PNR). We start with the general expression[5] of

$$\mathrm{PNR}^2=\left\{\frac{\langle I_c(0)\rangle}{\frac{1}{B_x-\Delta_x}\left[\int_{-B_x/2}^{B_x/2}\langle I_c(x)\rangle\,dx-\int_{-\Delta_x/2}^{\Delta_x/2}\langle I_c(x)\rangle\,dx\right]}\right\}, \tag{18}$$

where the denominator represents the root-mean-square amplitude across a correlation plane of spatial bandwidth B_x, which excludes a small region of width Δ_x centered around the correlation peak. With the filter plane SLM modeled as an array of equally spaced pixels of pitch Δ_f, the nonredundant bandwidth is $B_x=\Delta_f^{-1}$; i.e., the spacing between diffraction orders of the array. The width $\Delta_x=B_x/N$ corresponds to the resolution of an N-element array and these definitions also lead to $B_f=N\Delta_f=\Delta_x^{-1}$. These choices have been made specifically so that our model closely corresponds to typical fast-Fourier-transform-based computer simulations that represent each filter plane pixel with one sample. Consistent with these assumptions, we approximate the second integral in the denominator of Eq. (18) as $\Delta_x\langle I_c(0)\rangle$. The first integral is the energy in the central diffraction order. It can be related to the energy in the filter plane through Parseval's theorem. The total energy in the filter plane and the correlation plane is

$$E_f=\int_{-B_f/2}^{B_f/2}a^2(f)\,df=B_f\overline{a^2}=E_c . \tag{19}$$

The energy in the central diffraction order is found by recognizing that the ratio of energy in each diffraction order is determined from the Fourier series of a square wave grating of duty cycle $D=w/\Delta_f$, where w is the width of each pixel. Therefore the first integral in Eq. (18) can be written as D^2E_f. Using these results, Eq. (18) can be written as

$$\mathrm{PNR}=\left[\frac{(N-1)\langle I_c(0)\rangle}{N^2w^2\overline{a^2}-\langle I_c(0)\rangle}\right]^{1/2} . \tag{20}$$

Using Eq. (14) for $\langle I_c(0)\rangle$ in Eq. (20) completes the derivation of PNR (this expression is given presently). The PNR measured in an actual correlator will be subject to statistical fluctuations. One way to estimate the sensitivity of the measured PNR to randomness is to perturb $\langle I_c(0)\rangle$ the expected intensity in Eq. (20) by $\pm\sigma_I(0)$ from Eq. (16), and then use Eq. (14) to get

$$\mathrm{PNR}\pm\delta\mathrm{PNR}=\left\{\frac{(N-1)[pp_s+qZ/N\pm\sigma_I(0)/(Nw\overline{a})^2]}{Z-[pp_s+qZ/N\pm\sigma_I(0)/(Nw\overline{a})^2]}\right\}^{1/2} . \tag{21}$$

Setting the standard deviation to zero in Eq. (21) gives the expression for PNR. Equation (21) for the PNR and the perturbed PNR has no dependence on pixel width or duty cycle. [The factors of $Nw\overline{a}$ in Eq. (21) cancel with those in $\sigma_I(0)$; see Eq. (16).] Equation (21) also shows that PNR essentially depends on the square root of the number of pixels in the SLM. The effect on PNR of the term Z, which characterizes the spatial pattern of attenuation across the SLM, is nearly reciprocal to that of N. This relationship between N and Z is quite accurate if $N>>Z>>1$ and it can even be accurate for $Z=1$ if the phase errors are large [i.e., if the denominator term in brackets in Eq. (21) is much less than unity]. This relationship has led us to define $(N-1)/Z$ as an "effective number" of SLM pixels that are uniformly illuminated (this definition is discussed further in Refs. 3 and 8).

4 Example Analysis of Combined Systematic and Random Errors

The combined effect on PNR of nonunity slope (described by parameter k) and uniformly distributed random phase (described by the parameter k_r) is calculated using Eq. (21) [together with the definitions in Eqs. (15) to (17)]. The analysis is for a 128×128 SLM, i.e., $N=16{,}384$ and a transmitted amplitude spectrum of $Z=6.23$. These values were chosen to enable comparisons with the results given in Ref. 5 on systematic errors. The calculated results are presented in Figs. 1 through 5.

In Fig. 1, the curve for $k_r=1$ is identical to the modeled results for the case of nonunity slope in Ref. 5. For large values of k the other curves are depressed by a factor of roughly $p^{1/2}$. This follows from the term Z being much larger than pp_s and pp_s being much larger than qZ/N for large values

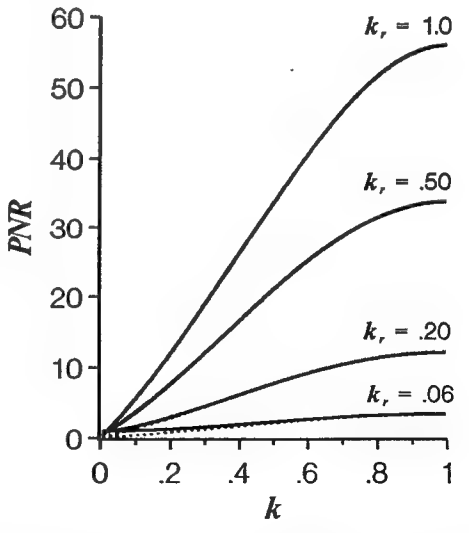


Fig. 1 PNR as a function k for various values of k_r; PNR is plotted as solid lines and its approximation as dotted lines.

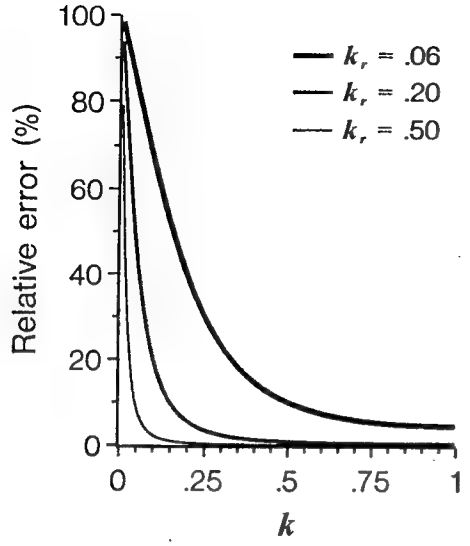


Fig. 2 Relative error between PNR and the approximation to PNR plotted as a function of k and for the same values of k_r as in Fig. 1.

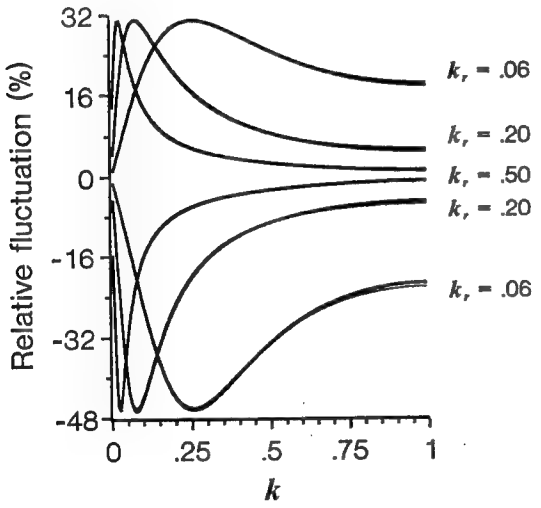


Fig. 3 Fluctuation in PNR resulting from perturbation of intensity by plus and minus the standard deviation of the expected intensity plotted as a function of k and for the same values of k_r as in Figs. 1 and 2; fluctuation is shown as heavy lines and approximations to fluctuations are shown as light lines.

of k. The term qZ/N can be practically ignored in this example, because the dashed curves that represent the values of PNR calculated with the term set to zero are only distinguishable at the lowest levels of PNR. In Fig. 2, the difference between PNR and its approximation is plotted as relative error with respect to PNR. The largest relative error is for $k=0$. At this point, PNR equals one and its approximation equals zero.

Figure 3 shows $\pm\delta$PNR/PNR the relative fluctuation for the PNR curves in Fig. 1. The smallest fluctuations are found for values of k and k_r near unity. As the values of these parameters decrease, the fluctuation increases, but only to a point. Below that point the fluctuation decreases. The reason for the decrease is that the intensity of the correlation peak is being dominated by noise or speckle background. The noise-dominated region occurs when pp_s [which is proportional to $|\langle c(0)\rangle|^2$] is less than qZ/N (which is proportional to the average intensity of the noise background). The turning point in each of the curves occurs when pp_s is of the same order of magnitude as qZ/N. The light lines in Fig. 3 are approximations in which all but the first of the four terms inside the braces in Eq. (16) have been omitted in the calculations. The differences are barely distinguishable and are primarily due to the second of the four terms. (We originally used values of $Z_3=Z^{1.5}$ and $Z_4=Z^2$, but numbers a few orders of magnitude larger cause no noticeable difference). What is interesting is that correlation performance can be predicted with such a small amount of specific information about the amplitude spectrum of the signal. It is also interesting that the percentage fluctuation in PNR is quite low for substantial amounts of random and systematic phase errors. This is a result of the coherent superposition of wavefronts from the typically large number of pixels in current SLMs.

An alternate way to plot Eq. (21) is to define k_r in terms of k. In particular, we have chosen to plot PNR for a constant ratio of phase mismatch

$$\alpha=\frac{1-k_r}{1-k}. \tag{22}$$

That is, as the slope of the phase error $1-k$ increases, the random phase spread $\nu=2\pi(1-k_r)$ increases proportionally. This is plotted in Fig. 4 for various values of α. For instance, for the case $\alpha=k_r/k=1$, PNR is reduced over the nonrandom case by no more than 40% for k greater than 0.5. The corresponding fluctuation curve in Fig. 5 shows that random fluctuation resulting from one-standard-deviation perturbation in the expected intensity of the correlation peak is less than around 3% for k greater than 0.5. To relate the magnitude

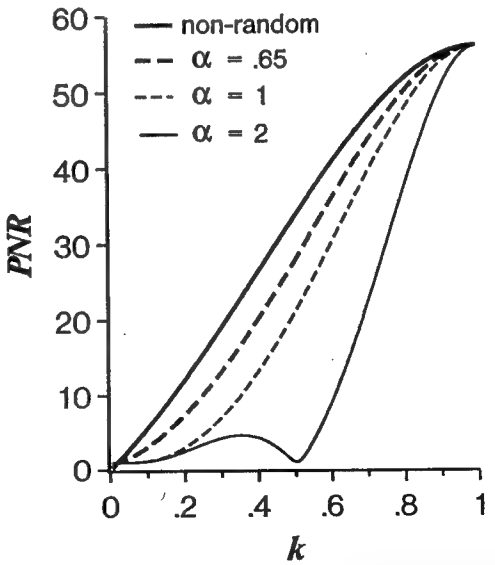


Fig. 4 PNR shown as a function of k for various values of α, a constant proportional to the ratio of random to systematic phase mismatch.

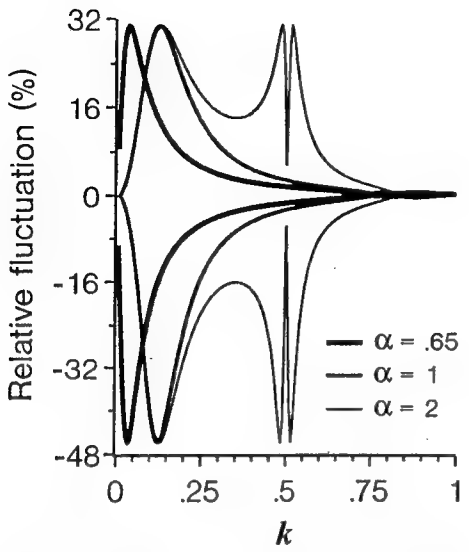


Fig. 5 Relative fluctuation of PNR resulting from perturbation of the expected intensity by plus and minus the standard deviation of intensity plotted as a function of k for the same values of α as in Fig. 4.

of these effects to well-known optical correlators note that for $k=0.5$ the model also describes the systematic phase errors produced by a binary phase-only filter (see Table 1). The PNR for the binary filter is reduced by roughly 40% over that of the analog phase-only filter. Phase randomness with a uniform spread over π (i.e., $\alpha=1$) reduces the PNR for the binary case by another 40%. The random fluctuations are also useful in establishing the maximum level for the decision threshold. For example, for this noise-corrupted correlator, dropout or misclassification of the correlation peak is highly unlikely for thresholds set 10% below PNR [corresponding to a perturbation from the expected peak intensity of roughly $-3\sigma_I(0)$].

Also note in Fig. 5 that the fluctuation is essentially zero for the $\alpha=2$ curve at $k=0.5$. This is not a noise-immune operating point. Instead, PNR has dropped to a value of one and thus (as discussed for Fig. 3) the expected correlation peak intensity is small compared to the average intensity of the noise/speckle background.

5 Conclusions

We have developed a model that describes the performance of optical correlators subject to a combination of phase errors. Various types of phase errors are related through a single parameter k. Their effects on peak correlation amplitude are given in Tables 1 and 2. The effect on correlation amplitude of other systematic and random phase errors of interest can also be evaluated [using Eq. (8)], and in many cases, reduced to a functional form. This result can then be inserted in the model for PNR [Eq. (21) with the standard deviation set to zero] and the statistical fluctuation of PNR [Eq. (21) along with evaluation of the standard deviation using Eqs. (16) and (17)].

It was possible to develop such a simple method by decoupling any dependence between the phases and the amplitudes in the filter plane. When this is done, the influence of the transmitted amplitude spectrum is totally contained in the single parameter Z, which is the normalized second moment of the amplitude spectrum. This approximate model provides much insight and is useful at the early stages of design. It should be considered prior to performing exhaustive simulations and design studies on the effects of various phase errors. As part of this analysis, it may be desirable to consider the values of Z for the objects in the image training set.

Of course, for many SLMs the values of phase and amplitude are coupled.[10] We did not consider this problem here. It seems reasonable, however, that there are alternate evaluations or approximations of Eq. (6) that can lead to models of nearly comparable usefulness and simplicity.

Acknowledgments

Discussions with J. L. Horner, Rome Laboratory, and W. A. Friday, U.S. Army Missile Command, stimulated this line of investigation and are gratefully appreciated. Detailed comments by the reviewers are appreciated. One reviewer also deserves credit for first noting that this analysis is insensitive to constant phase errors. The study was funded by the Advanced Research Projects Agency (ARPA) through U.S. Air Force Rome Laboratory contract No. F19628-92-K-0021 and NASA through Western Kentucky University under cooperative agreement NCCW-60.

References

1. J. C. Dainty, Ed., *Laser Speckle and Related Phenomena*, 2nd ed., Springer, Berlin (1984).
2. J. L. Horner and P. D. Gianino, "Effects of quadratic phase distortion on correlator performance," *Appl. Opt.* **31**(20), 3876–3878 (1992).
3. R. W. Cohn and J. L. Horner, "Limited phase modulation and its effect on phase-only correlation," *Proc. SPIE* **2237**, 147–151 (1994).
4. W. J. Smith, *Modern Optical Engineering*, Ch. 11, McGraw-Hill, New York (1966).
5. R. W. Cohn and J. L. Horner, "Effects of systematic phase errors on phase-only correlation," *Appl. Opt.* **33**(23), 5432–5439 (10 August 1994).
6. R. W. Cohn and R. J. Nonnenkamp, "Statistical moments of the transmittance of phase-only spatial light modulators," *Proc. SPIE* **1751**, 289–297 (1992).
7. R. W. Cohn, "Random phase errors and pseudorandom phase modulation of deformable mirror spatial light modulators," *Proc. SPIE* **1772**, 360–368 (1992).
8. R. W. Cohn and M. Liang, "Approximating fully complex spatial modulation with pseudorandom phase-only modulation," *Appl. Opt.* **33**(20), 4406–4415 (10 July 1994).
9. B. V. K. Vijaya Kumar and L. G. Hassebrook, "Performance measures for correlation filters," *Appl. Opt.* **29**(20), 2997–3006 (1990).
10. R. D. Juday, "Correlation with a spatial light modulator having phase and amplitude cross coupling," *Appl. Opt.* **28**(22), 4865–4869 (1989).
11. J. L. Horner, "Metrics for assessing pattern-recognition performance," *Appl. Opt.* **31**(2), 165–166 (1992).

Robert W. Cohn received the PhD in electrical engineering from Southern Methodist University, Dallas, Texas, in 1988. He joined the University of Louisville as associate professor of electrical engineering in 1989. From 1978 to 1989 he was with Texas Instruments, Inc., as a member of the technical staff from 1984 to 1989 in the Central Research Laboratories and the Defense Systems Division, where his research included compensation of second-order effects in surface acoustic wave (SAW) filters, optical processing using deformable mirrors, and image processing algorithms for guidance and tracking. As a design engineer from 1978 to 1983 in Defense Systems he investigated SAW filter compensation methods and developed microwave hybrid components. Cohn is a senior member of IEEE and a member of SPIE and OSA.

a reprint from Applied Optics

Approximating fully complex spatial modulation with pseudorandom phase-only modulation

Robert W. Cohn and Minhua Liang

Any desired diffraction pattern can be produced in the Fourier plane by the specification of a corresponding input-plane transparency. Complex-valued transmittance is generally required, but in practice phase-only transmittance is used. Many design procedures use numerically intensive, constrained optimization. We instead introduce a noniterative procedure that directly translates the desired but unavailable complex transparency into an appropriate phase transparency. At each pixel the value of phase is pseudorandomly selected from a random distribution whose standard deviation is specified by the desired amplitude. We also derive statistical expressions and use them to evaluate the approximation errors between the desired and achieved diffraction patterns.

Key words: Optical information processing, spatial light modulators, phase-only filters, laser speckle, rough surface scattering, statistical optics, binary and diffractive optics, phased arrays.

Introduction

This study is motivated by our desire to design phase-only filters and diffractive elements with a small amount of electronic computation, and thereby permit programming of arbitrary spatial modulation at real-time rates. Popular design procedures (e.g., the Dammann grating,[1–3] simulated annealing,[4] iterative constrained optimization,[5,6] and other iterative procedures) are only practical if performed off line because of both the numerical cost of performing Fourier transforms repeatedly and the further cost of evaluating the sensitivity of the transform with respect to a large number of pixels (frequently every pixel of the input-plane spatial light modulator.) Although it can be argued that for some applications all necessary phase-only filters can be computed off line and stored in memory, in other applications either too much memory is needed to do this affordably or there may not be advance knowledge of what filter is needed.

There are many procedures in the area of computer-generated holography, especially kinoforms, that permit direct synthesis of the input plane. These presuppose that the Fourier transform pair between the fully complex-valued input and the Fourier planes is known, and they work by encoding the desired complex values to appropriate phase settings. The direct-synthesis design procedures thus permit programming at real-time rates if the desired Fourier plane pattern is known.[7] The amount of memory is also minimized if the complex-valued Fourier transform pair can be written as an easily computed function.

Most frequently the direct procedures group a few adjacent pixels together as a single pixel that approximates several discrete settings over the complex plane (i.e., cell-oriented encoding).[8] However, this reduces the space bandwidth, which is already quite small (say 128 × 128 pixels) for current spatial light modulators, as compared with traditional fixed pattern holographic and diffractive optical elements. The procedure we present here is also a direct method, but one for which a continuous value of phase is selected for each individual pixel, independent of all other settings (i.e., point-oriented encoding.)[8]

Our method directly follows from mathematical models (presented below) of far-field diffraction from arrays of randomly phased point sources. More specifically, the phases are treated as independent and *nonidentically* distributed random variables. Many previous studies have considered random phase diffusers in which the phase statistics are identically distributed across the diffuser.[9,10] Their objective was improved holographic reconstruction of the intensity of objects, for which the phase was of no concern to the observer. In our design we are interested in reconstructing full complex objects from a variably random phase-only filter plane. In our analysis and

The authors are with the Department of Electrical Engineering, University of Louisville, Louisville, Kentucky 40292.

Received 15 September 1993; revised manuscript received 23 December 1993.

0003-6935/94/204406-10$06.00/0.

design procedure it is convenient to represent the random variables with a single probability-density function that is varied in its mean and standard deviation. Our analysis shows that, on average, the pixels with phases drawn from distributions of larger standard deviation behave as if they have smaller amplitude transmittances. This result originally occurred to us when we noted an average contrast loss for arrays of phase-only pixels that have identically distributed phase errors.[11] We now recognize that this earlier result is just a special case of the more general result for nonidentically distributed phases.

The ideal spatial light modulator (SLM) for our procedure modulates phase continuously over a full 2π range. Regularly spaced pixels are assumed in the examples below but are not required in our theory. These characteristics are typical of those anticipated for the piston-modulating (or flexure-beam) deformable mirror device[12] and are achievable with birefringent[13] and twisted nematic[14] liquid crystals. For binary and multilevel optics the quantized values of phase can be modeled as statistical departures from the desired analog phase, and this added effect on the diffraction pattern can be estimated.[11,15]

Concept

Our design procedure specifies the degree of randomness (i.e., standard deviation) of phase at each pixel as a way of approximating arbitrary values of amplitude. This is analogous to placing a diffuser of spatially varying roughness over the input plane. Increasing roughness at a pixel decreases its coherent contribution to the far-field pattern. The remaining incoherent or diffused light is spread over the entire diffraction pattern and contributes a noise background that is frequently referred to as speckle. The design phase at any pixel is specified as the expected value of its random phase distribution. Thus the full complex input plane can be viewed as a cascade of a deterministic phase screen with a variable roughness phase screen.

This interpretation follows from the statistical expectation of a complex exponential of a random argument, which is schematically illustrated in Fig. 1 and is derived in the following section. Figure 1 shows the spatial phase and amplitude for two types

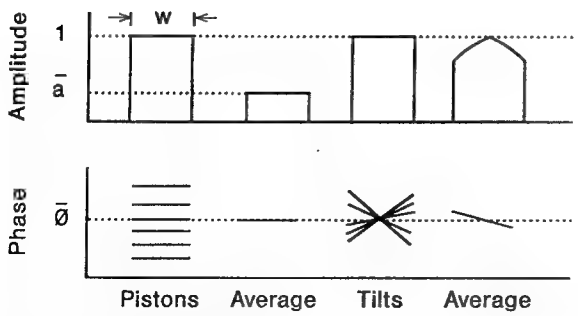


Fig. 1. Expected transmittance of random phase-only pixels (both piston-only and tilt-only pixels).

of phase-only pixels: piston-only pixels, those that produce a single value of phase across the pixel width, and tilt-only pixels, those that produce a linearly varying value of phase across the pixel width. The design procedure is concerned with the former structure, and the latter is provided as a further example of the concept. The leftmost of the four plots of amplitude and phase represents a random phase pixel. Because phase is a random variable we show an ensemble of phase values. The ensemble average of the random complex exponential produces the result in the second column: an average value of phase and a loss in amplitude transmittance. The example of pixels with random tilts shows a similar result in that the average amplitude transmittance decreases with increasing phase fluctuations that directly correspond with distance from the pivot point. Further discussion and analysis of the effects of random tilts are described in Ref. 11.

Because both expectation and Fourier transform are linear operations, the expected complex amplitude of the far-field pattern of the piston-only pixel, as well as an array of pixels, is the Fourier transform of the expected transmittance. The expected intensity of the far-field pattern consists of the magnitude squared of the complex amplitude plus a broad pedestal caused by the average intensity of speckle.

In our design procedure, rather than average many trials, we instead select a single value of phase from the ensemble for each pixel. The expected complex amplitude and intensity of the far-field pattern for this situation is mathematically equivalent to that described in the previous paragraph. More importantly, the actual far-field pattern approximately resembles the average pattern that results from the coherent summation of a large number of random wave fronts. Thus the formation of the far-field pattern can be viewed as a physical example of the central limit theorem, i.e., the so-called law of large numbers.[16]

Mathematical Analysis

Deterministic Expressions

The following definitions for arrays of piston-only pixels are used in the development of the statistical expressions. The complex transmittance of an individual pixel centered at position x equal to x_i in the input plane will be written as $a_i(x)$, and the transmittance of the array of the N individual pixels is

$$t(x) = \sum_{i=1}^{N} a_i(x) = \sum_{i} r(x - x_i)\exp(j\psi_i), \qquad (1)$$

where ψ_i is the phase shift produced by the ith pixel of the SLM. The appreviation $r(x) = \mathrm{rect}(x/w)$ has been used where w is the width of each pixel. Equation (1) shows that the amplitude and phase of each pixel have been defined in local coordinates centered around $x = 0$ and then shifted to pixel locations x_i. In this and subsequent equations the

inactive area between the pixels is treated as being zero amplitude. We do this to focus on the fundamental effect. It is also a reasonable assumption in light of current SLM's that either have nearly unity fill factors (in the case of deformable-mirror devices and binary optics) or that are fabricated with opaque shadow masks (such as current liquid-crystal displays.) The intensity of the far-field diffraction pattern is written

$$I(f_x) = T(f_x)T^*(f_x) = \mathcal{F}\{t(x) \oplus t(x)\}, \quad (2)$$

where $T(f_x)$ is the Fourier transform of the transmittance $t(x)$, i.e. $T(f_x) = \mathcal{F}\{t(x)\}$, and where $\oplus$ indicates the correlation integral

$$a(x) \oplus b(x) = \int a(x' + x)b^*(x')\mathrm{d}x'. \quad (3)$$

The squared intensity spectrum expressed in terms of the fourth-order autocorrelation of SLM transmittance is

$$I^2 = TT^*T^*T = \mathcal{F}\{[t(x) \oplus t(x)] \oplus [t(x) \oplus t(x)]\}. \quad (4)$$

Unlike convolution, the order in which correlations are performed affects the result, and thus the brackets are required in Eq. (4).

Statistical Expressions

The expected value of a complex phasor of a random argument is frequently referred to as the characteristic function[16] of the random variable ψ:

$$M'(\omega) = \langle \exp(j\omega\psi) \rangle = \int_{-\infty}^{\infty} \exp(j\omega\psi)p_\psi(\psi)\mathrm{d}\psi$$

$$= 2\pi\mathcal{F}^{-1}\{p_\psi(\psi)\}, \quad (5)$$

where $\langle\ \rangle$ is the ensemble average (i.e., expectation) operator and where $p_\psi(\psi)$ is the probability-density function of ψ. The last equality in Eq. (5) emphasizes the well-known Fourier transform relationship between the probability of the density function and the characteristic function. If the phases ψ_i are independent random variables, then the expected value of pixel transmittance $a_i(x)$ is

$$\bar{a}_i(x) = \langle a_i(x) \rangle = \langle r(x - x_i)\exp(j\psi_i) \rangle,$$
$$= r(x - x_i)M_i'(1),$$
$$= r(x - x_i)\exp(j\bar{\psi}_i)M_i(1), \quad (6)$$

where M_i' is the characteristic function of ψ_i and M_i is the characteristic function of the unbiased values of phase $\psi_i - \langle \psi_i \rangle$.[17] We have written the last line of Eq. (6) to identify the magnitude and phase components, and it corresponds to the average piston case in Fig. 1 (second column).

We considered two specific probability distributions for phase ψ_i. For Gaussian distributed phase of standard deviation σ_i,

$$M_i(1) = \exp\left(-\frac{1}{2}\sigma_i^2\right), \quad (7)$$

and for uniformly distributed phase with total spread,

$$v_i = \sqrt{12}\sigma_i, \quad (8)$$

$$M_i(1) = \mathrm{sinc}\left[\frac{v_i}{2\pi}\right]. \quad (9)$$

For either distribution the average amplitude transmittance at each pixel i can be controlled explicitly by selection of the value of the standard deviation σ_i. In digital simulations we prefer the uniform distribution because most Gaussian random-number generators are derived by performing additional numerical operations on uniform samples. A further advantage of using uniform statistics is that the total spread never need exceed 2π, whereas with Gaussian statistics the standard deviation can be infinite, and which can lead to overflow and underflow errors if not handled carefully.

The expected value of Eq. (1), using either probability distribution, is the expected SLM tranasmittance

$$\bar{t}(x) = \sum_{i=1}^{N} \langle a_i(x) \rangle = \sum_i r(x - x_i)p_i^{1/2}\exp(j\bar{\psi}_i), \quad (10)$$

where we use the parameter

$$p_i = M_i^2(1) \quad (11)$$

to simplify subsequent expressions. Note that because the expectation and Fourier transform operators are both linear, the expected far-field transmittance (more precisely, the angular spectrum) is

$$\bar{T}(f_x) = \mathcal{F}\{\bar{t}(x)\}. \quad (12)$$

The nomenclature in Eq. (12) of using lowercase letters for space-domain variables and uppercase variables for frequency-domain variables will be used throughout this discussion.

The most general expectation for the intensity of the far-field diffraction pattern of SLM's with statistically independent pixels follows from the first equality in Eqs. (1) and (2). It is

$$\bar{I}(f_x) = \sum_i \sum_j \langle A_i(f_x)A_j^*(f_x) \rangle,$$
$$= \sum_{i \neq j} \sum \langle A_i \rangle \langle A_j^* \rangle + \sum_i \langle |A_i|^2 \rangle,$$
$$= \sum_i \sum_j \langle A_i \rangle \langle A_j^* \rangle - \sum_i |\langle A_i \rangle|^2 + \sum_i \langle |A_i|^2 \rangle,$$
$$= |\bar{T}|^2 + \sum_i [\langle |A_i|^2 \rangle - |\bar{A}_i|^2], \quad (13)$$

where the intermediate results indicate how we used independence to simplify the expression. This expression shows that the expectation nearly separates into $|\langle T\rangle|^2$, except for terms involving the autocorrelation of the input-plane pixels (i.e., terms for which $i = j$). Under the same assumption of independent pixels, we find that the most general expression for the squared far-field intensity is

$$\begin{aligned}\langle I^2\rangle &= \sum_i\sum_j\sum_k\sum_l \langle A_iA_j^*A_k^*A_l\rangle \\ &= 2[\bar{I}^2 - |\bar{T}|^4] + \left|\bar{T}^2 + \sum_i(\langle A_i^2\rangle - \bar{A}_i^2)\right|^2 \\ &\quad + 4\,\mathrm{Re}\Big[\bar{T}^*\sum_i(\langle|A_i|^2A_i\rangle - \langle A_i^2\rangle\bar{A}_i^* \\ &\quad + 2|\bar{A}_i|^2\bar{A}_i - 2\langle|A_i|^2\rangle\bar{A}_i)\Big] \\ &\quad + \sum_i[\langle|A_i|^4\rangle - 6|\bar{A}_i|^4 + 8\langle|A_i|^2\rangle|\bar{A}_i|^2 \\ &\quad - |\langle A_i^2\rangle|^2 - 2\langle|A_i|^2\rangle^2 \\ &\quad + 4\,\mathrm{Re}(\langle A_i^2\rangle\bar{A}_i^{*2} - \langle|A_i|^2A_i\rangle\bar{A}_i^*)]. \end{aligned} \tag{14}$$

It was found by substituting Eq. (1) into Eq. (4) and then taking the expectation of Eq. (4). The second quality was found by a procedure similar to that of Goodman.[18] (See Appendix A for more details.) The standard deviation of intensity $\sigma_I(f_x)$ is then directly found through the use of the well-known result

$$\sigma_I^2 = \langle I^2\rangle - \bar{I}^2. \tag{15}$$

For the specific case of piston-only, nonidentically distributed random phase, Eqs. (13)–(15) simplify to

$$\begin{aligned}\bar{I}(f_x) &= \mathcal{F}\Big\{\bar{t}(x)\oplus\bar{t}(x) + [r(x)\oplus r(x)]\sum_i q_i\Big\}, \\ &= |\bar{T}(f_x)|^2 + R^2(f_x)\sum_i q_i, \end{aligned} \tag{16}$$

where

$$q_i = 1 - p_i, \tag{17}$$

$$\begin{aligned}\sigma^2(f_x) &= \bar{I}^2 - 2|\bar{T}|^4 + |\bar{T}^2 - T_A|^2 \\ &\quad - 4\,\mathrm{Re}[\bar{T}T_B^*] - G_A. \end{aligned} \tag{18}$$

We introduce a shorthand for defining repeated autocorrelations of the rect function $r(x)$:

$$g_n(x) = [r(x)\oplus\cdots\oplus r(x)]_n, \tag{19}$$

where n indicates the number of rect functions (i.e., $n - 1$ integrals). Equation (18) is completely specified with the additional definitions for the Gaussian distribution

$$\begin{aligned} t_A(x) &= \sum_i g_2(x - 2x_i)\exp(j2\bar{\psi}_i)q_ip_i, \\ t_B(x) &= \sum_i g_3(x - x_i)\exp(j\bar{\psi}_i)q_i^2p_i^{1/2}, \\ g_A(x) &= g_4(x)\sum_i q_i^4, \end{aligned} \tag{20}$$

and for the uniform distribution

$$\begin{aligned} t_A(x) &= \sum_i g_2(x - 2x_i)\exp(j2\bar{\psi}_i)(p_i - \mathrm{d}_i), \\ t_B(x) &= \sum_i g_3(x - x_i)\exp(j\bar{\psi}_i)(q_i - p_i + \mathrm{d}_i)p_i^{1/2}, \\ g_A(x) &= g_4(x)\sum_i (q_i - 3p_i + 6p_i^2 - 4p_i\mathrm{d}_i + \mathrm{d}_i^2), \end{aligned} \tag{21}$$

where

$$\mathrm{d}_i = \mathrm{sinc}\left[\frac{v_i}{\pi}\right]. \tag{22}$$

Discussion and Interpretation

The expected intensity pattern, Eq. (16), contains the desired design intensity pattern, the magnitude square of Eq. (12), plus an additional term proportional to $G_2(f_x) = R^2(f_x)$ that we associate with the average level of speckle intensity, and which is often referred to as halo or pedestal.[19] This second term is proportional to the far-field diffraction pattern intensity of a single pixel. Examination of Eq. (18) also reveals that each of its terms has the common factor $G_4(f_x) = G_2^2(f_x)$. Thus for a standard definition of signal-to-noise ratio (SNR),

$$\mathrm{SNR}(f_x) = \frac{\bar{I}(f_x)}{\sigma_I(f_x)}, \tag{23}$$

the term $G_2(f_x)$ cancels out and is independent of spatial frequency. The reciprocal of Eq. (23) basically describes the relative approximation error (especially when σ_I is small compared with I and the contribution of pedestal can be practically ignored).

Another type of SNR that is quite common in describing the quality of an optical correlation peak is the ratio of peak intensity to background noise level; this is often referred to as the peak-to-noise ratio (PNR). The analysis above permits calculation of this as well; however, unless the pixels are point sources, the pixel element factor $G_2(f_x)$ has to be considered and so there will be different values of the SNR depending on the pixel fill factor.

Summary of the Design Procedure

The above analysis indicates that a phase-only source distribution can be treated as a full complex distribution for the purpose of approximating a desired

far-field pattern. The results from above can be used in a design procedure as follows. Specify the desired far-field pattern $T(f_x)$. Fourier transform to the desired source distribution $t(x)$. For each pixel i, invert Eq. (9) [preferred, or Eq. (7)] to find σ_i, and select a pseudorandom number, appropriately scaled by σ_i to represent the random phase ψ_i. Then Fourier transform the array of random phase-only pixels to evaluate the actual far-field pattern. If desired, compare the actual far-field pattern $I(f_x)$ with Eqs. (16), (18), and (23). In the next section we use this procedure to illustrate designs of diffraction patterns, including evaluations of the quality of the approximation.

Computer Verification and Demonstration of the Theory

General Experimental Conditions

Several designs have been simulated and analyzed, two of which are presented in some detail here. In every case the design is for a 128 × 128 pixel phase-only SLM. It is represented as a 128 × 128 array of samples embedded in the center of a 512 × 512 array of zeroes. We use a fast Fourier transform (FFT) routine to approximate the Fourier transform of the array and of the individual terms comprising the expectations in Eqs. (16) and (18).

The uniform random-number generator DRNUNF (from the IMSL numerical software library from IMSL, Inc., Houston, Tex. 77042-3020) with initialization subroutine RNOPT(6) was used. We mention this because initially there were problems when we used the random-number generator RAN1 given in Ref. 20. Specifically, when we attempted to verify Eqs. (16) and (18) by Monte Carlo estimation (i.e., ensemble averaging of a large number of identical designs generated with different random-number seeds), the estimates sometimes converged to different results than in the equations. The discrepancies were especially obvious at harmonically related sets of spatial frequencies. This may indicate that RAN1 produced correlated sequences in these experiments. These problems, however, were not discernible in individual design runs because of the magnitude of the random fluctuations.

When using DRNUNF we did find that plots of Monte Carlo estimates of the expected value and standard deviation of intensity, except for small fluctuations, appear to match our expressions closely when 1000 design runs are used in the estimate. We also checked that the relative error between the theory and the estimate decreases with an increasing number of design runs and that this improvement is roughly proportional to the square foot of the number of design runs used in the estimate, as would be expected. Our closest result, using 10^5 design runs, converged to the theoretical expressions with a fluctuation of less than 0.4% standard deviation.

Pseudorandom Encoding of Binary Amplitudes

Figures 2 and 3 illustrate the design of a phase modulation to approximate an elliptically shaped aperture of eccentricity 6:1. Figure 2(a) shows the desired amplitude modulation, and Fig. 2(b) shows the corresponding phase-only modulation. The mottled region in Fig. 2(b) corresponds to phase that is totally random (uniform statistics with spread $v_i = 2\pi$), and the white region corresponds to constant phase ($v_i = 0$). Figure 2(c) shows the expected intensity, Eq. (16) from the random modulated SLM, and Fig. 2(d) shows the resulting intensity pattern. It is basically an elliptical version of the Airy pattern (the exact far-field pattern of an elliptical aperture) on top of a uniform intensity (dark gray) background. The actual diffraction pattern [Fig. 2(d), the far-field pattern of the phase-only modulation] also resembles the elliptical Airy pattern on top of a speckle pattern. Figure 3 presents quantitative information along the vertical axis of Figs 2(c) and 2(d). The thin curves depict the error bars of $\pm\sigma_I$ added to the expected intensity. These provide a rough idea of the magnitude of the actual fluctuations, both of the coherent peak and the speckle-dominated sidelobe region.

It should be apparent that fluctuations and thus approximation errors will be smaller for those binary amplitude designs that use a greater number of nonrandom pixels. This relationship can be seen in Eq. (23), which simplifies to

$$\mathrm{SNR}(0) = \frac{N_N^2 + N_R}{(2N_R N_N^2 - N_R + N_R^2)^{1/2}} \approx \frac{N_N}{\sqrt{2N_R}} \quad (24)$$

at zero frequency where the diffraction intensity peaks.[21] Here N_R is the number of randomly modulated pixels on the SLM and N_N is the number of nonrandomly modulated pixels, and their sum is the total number of SLM pixels, 16,384. The approximation in Eq. (23) is valid except for SNR close to unity; for instance, for 1000 nonrandom pixels the relative error is roughly 1% and the SNR is 5.7. Figure 4 plots relation (24) against 500 run Monte Carlo estimates of SNR(0). These estimates correspond to elliptical aperture designs, with a major axis of 128 pixels and eccentricity ranging from 1 to 13, and circular apertures of diameter from 42 to 128 pixels. For reference, there are 2016 nonrandom pixels in the ellipse of eccentricity 6:1. This data point on Fig. 4 is nearly indistinguishable from the point for a circular aperture with a diameter of 50 pixels and that contains 1976 pixels.

A second performance measure describing the quality of the far-field diffraction pattern is the PNR,[22–24] which we choose to define here as

$$\mathrm{PNR} \equiv \frac{\bar{I}(0)}{\sigma_I(f_x)} \approx \frac{N_N^2}{N_R} \approx 2\mathrm{SNR}^2(0) \approx \frac{\bar{I}(0)}{\bar{I}(f_x)}, \quad (25)$$

where f_x is assumed to be a frequency in the sidelobe region. The approximations follow by assuming the expected transmittance is negligible, i.e. $\langle T(f_x)\rangle = 0$ in Eqs. (16) and (18), with respect to the speckle intensity in the sidelobe region. Additionally, we set the ratio of $G_2(0)/G_2(f_x)$ to unity, mainly to indicate more

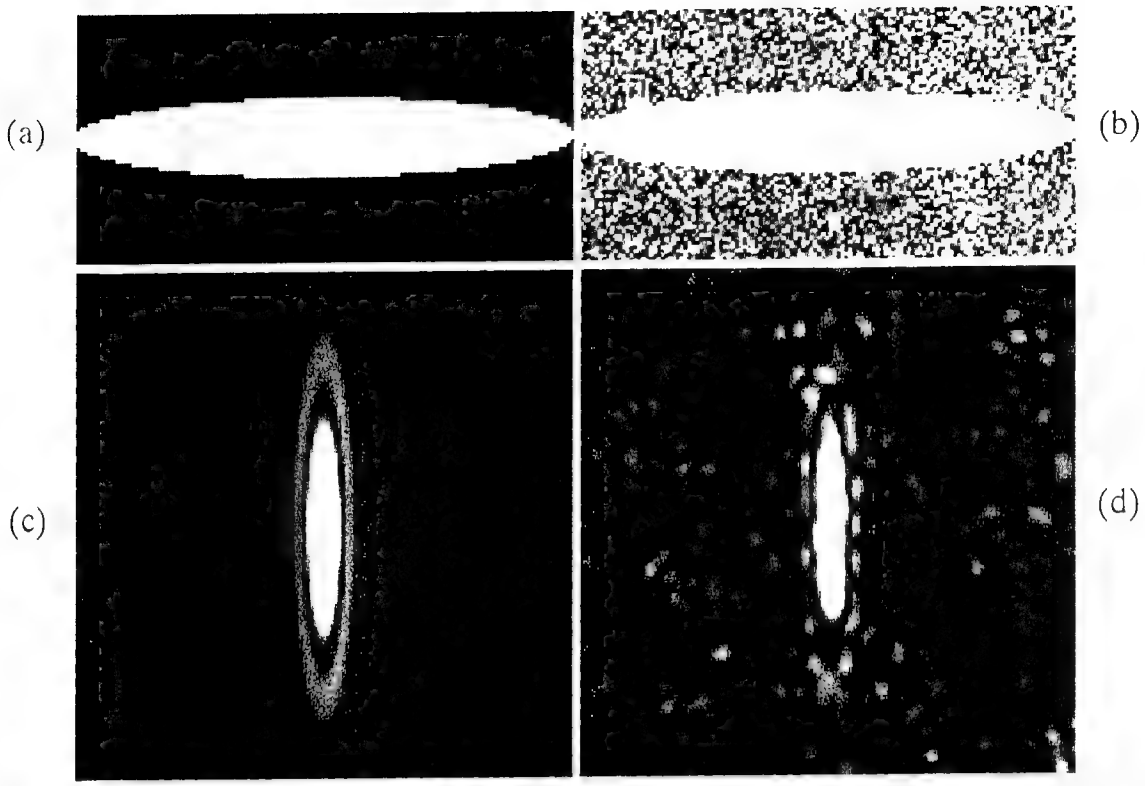


Fig. 2. Pseudorandom phase-only design of elliptical aperture. (a) Desired amplitude modulation. (b) Phase-only approximation to (a). (c) Expected value of (d) the intensity of the far-field diffraction pattern. Only the central 128 × 64 pixels of the 128 × 128 modulation are shown in (a) and (b). Only the central 128 × 128 pixels of the 512 × 512 FFT are shown in (c) and (d). We have nonlinearly transformed the intensities in (c) and (d) by the exponent 1.3 (i.e., gamma) to increase the contrast of low-lying sidelobes of the Airy pattern.

dramatically the simple relationship between PNR and SNR(0). In our simulations, for which we chose f_x as the Nyquist frequency [i.e., the point halfway between the (0, 0) and the (1, 1) diffraction order], the

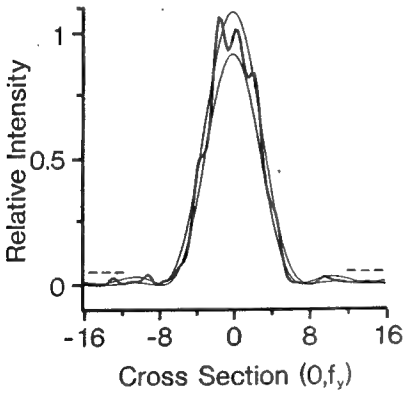


Fig. 3. Diffraction pattern of phase-only approximated elliptical aperture. The actual pattern (thick curve) is compared with the expected intensity plus and minus the error limits of one standard deviation. Horizontal dashed lines indicate the saturation (full-white) level of corresponding figures, Figs. 2(c) and 2(d). The spatial coordinates are normalized so that one unit corresponds to one of the 128 resolvable scan positions of the SLM or four samples of the FFT window

first approximation is within 1.5% of the exact value of the PNR for 1000 or more nonrandom pixels. The third approximation indicates that the average level of speckle in the sidelobe regions is equal to the standard deviation of intensity. This result is not unexpected based on previous observations that the intensity of fully developed speckle patterns is exponentially distributed and that the standard deviation of this distribution is equal to its mean.[11,19]

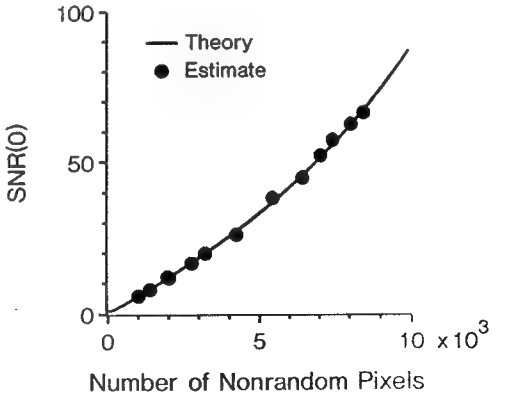


Fig. 4. SNR at diffraction peak as a function of number of nonrandom pixels for pseudorandom encoding of binary amplitudes.

We have also evaluated the design procedure for the approximation of an apodized input plane. Two-dimensional Gaussian, Airy pattern, and sinc functions with varying aspect ratios have been synthesized. We originally chose the sinc and Airy functions to produce brick-wall and top-hat diffraction patterns. However, the finite spatial extent of the SLM severely truncates the ideal functions and produces significant in-band ripple (Gibbs phenomena) and large sidelobes. The only feature that recommends this design is the extremely rapid transition band.

The transition bandwidth can be readily traded off with ripple and sidelobe level by the window design method that is widely used in the design of finite impulse response digital filters.[25] In this method one controls truncation effects by multiplying the ideal, infinite extent functions by an amplitude-tapered window in place of a rect function. One of the best windows for the design of uniform intensity patterns is the Dolph–Chebyshev function. It is considered optimal in the sense that its Fourier transform produces the minimum mainlobe width for a given sidelobe level.[26] The parameter α specifies the sidelobe level of 20α in decibels.

We have used the window method to design a pseudorandom phase modulation that will diffract into a close approximation of a brick-wall shape. The design function is a one-dimensional sinc function multiplied by a Dolph window in both coordinates. Our general design goal for this example is to produce a diffraction pattern with the largest aspect ratio (or eccentricity) possible that still reasonably resembles a brick-wall shape. We feel that for a 128×128 SLM this goal is met by the function $\mathrm{sinc}(4x)$ multiplied by Dolph functions in x and y that each have α equal to 1.3.

This result is presented in Figs. 5 and 6. In order to show the correspondence between amplitude in Fig. 5(a) and random phase in Fig. 5(b) more clearly, the gray-scale image shows, rather than phase error, the absolute phase error

$$\delta\psi(x_i) = |\psi_i - \bar{\psi}_i|. \tag{26}$$

The value of α has been chosen just large enough so that the sidelobes disappear in the speckle background in Fig. 5(d). They are still somewhat apparent in the expected intensity, Fig. 5(c). Figure 6(a) makes clear that increasing α further also increases the intensity of speckle, and overall passband ripple is not reduced because the random fluctuation is already larger than the ripple for the Dolph function. The aspect ratio of this diffraction pattern is 6.7:1 at

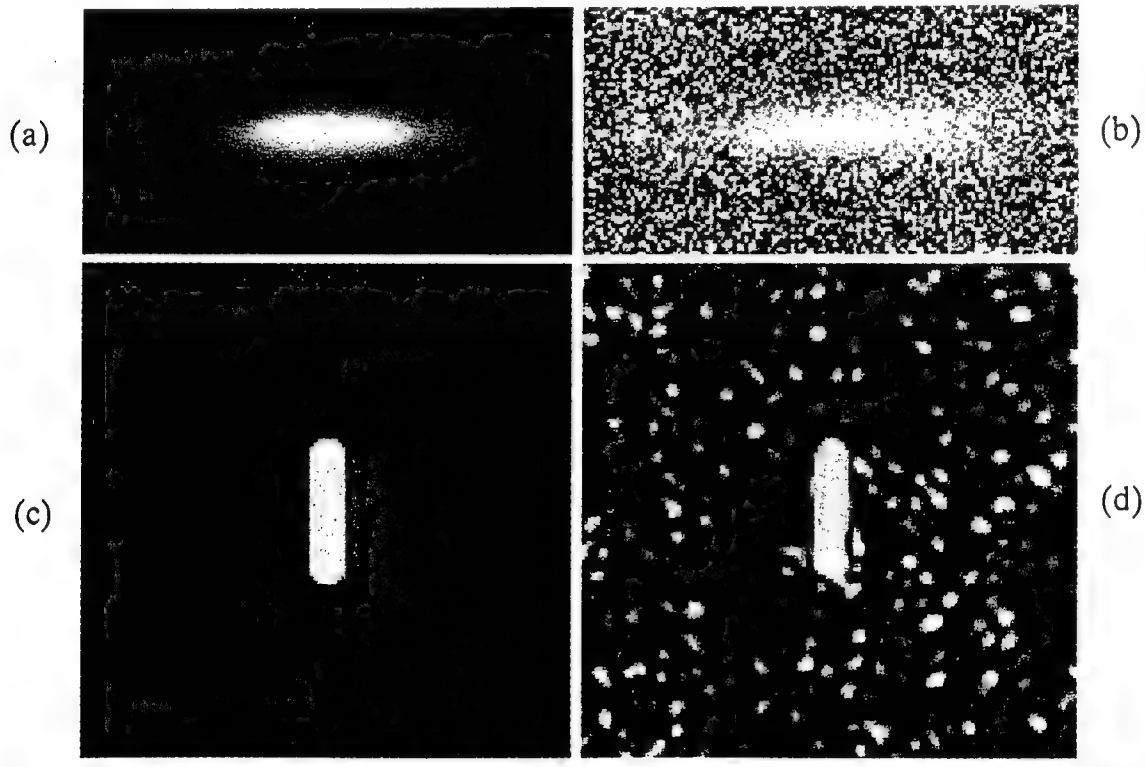


Fig. 5. Pseudorandom phase-only design of Dolph-windowed sinc apodization: (a) Desired amplitude modulation. (b) Phase-only approximation to (a). (c) Expected value of (d) the intensity of the far-field diffraction pattern. All units and settings are the same as in Fig. 2 except the recording gamma, which is unity for (c) and (d).

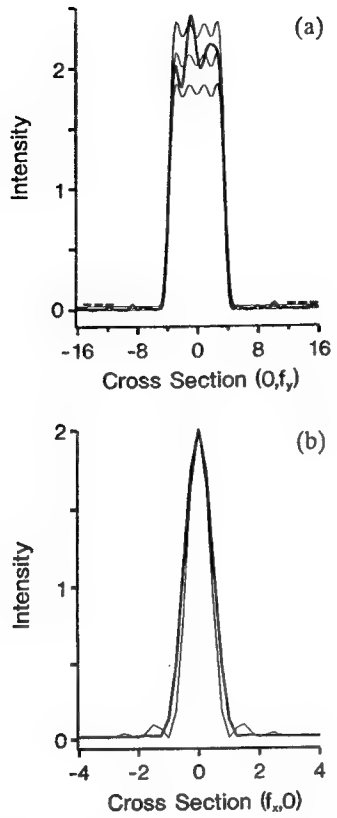


Fig. 6. Diffraction pattern of a phase-only approximated Dolph-windowed sinc apodization. (a) Cross section across vertical axis $(0, f_y)$. The actual pattern (thick curve) is compared with the expected intensity and the expected intensity plus and minus the errors limits of one standard deviation. Horizontal dashed lines indicate the saturation (full-white) level of corresponding figures, Figs. 5(c) and 5(d). Legend and units identical to those in Fig. 3. (b) Cross section across horizontal axis $(f_x, 0)$. The expected intensity (thick curve) is compared with the expected intensity $\text{sinc}^2(f_x)$ if no Dolph windowing had been employed.

the half-power points as measured from the plots in Fig. 6. Figure 6(b) also shows that the Dolph windowing in the vertical direction has widened the diffraction pattern by approximately 25% at the half-power points.

To compare the performance of various design functions with the phase-only encoding procedure, we introduce the definition of the effective number of random pixels,

$$\hat{N}_R \equiv \sum_i^N q_i. \tag{27}$$

Equation (27) is identically the amplitude of the noise pedestal term in Eq. (16), and it is identical to N_R for the case of binary modulation. Figure 7 compares the SNR for the various functions studied. The curves demonstrate a rough correspondence, with the apodized functions being more closely clustered than

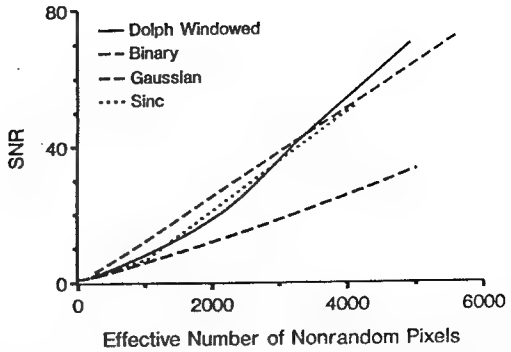


Fig. 7. SNR as a function of effective number of nonrandom pixels for pseudorandom encoding of continuous amplitudes. SNR(0) is plotted for Gaussian and binary, and average SNR across passband is plotted for Dolph and sinc designs. The sinc is a rectangularly symmetric function, and the Gaussian is a circularly symmetric function.

the binary amplitude functions. The circularly symmetric approximation to an Airy pattern on the input plane is not shown, but its performance nearly overlays that for the sinc curve in Fig. 7. For reference, the SNR of the design in Figs. 5 and 6 has a SNR of roughly 8, which corresponds to an effective number of 1000 nonrandom pixels.

Conclusions

We have presented a method, based on properties of random phase having spatially varying statistics, that approximates fully complex input-plane modulation. Diffraction patterns from pseudorandom phase-only modulation can be as energy efficient as any passive, fully complex modulation, with performance loss arising from the addition of a nearly uniform level speckle background. Designers can use the method to specify diffractive optical elements directly from Fourier transform relationships between the input plane and far field. The performance of any design can be readily evaluated, and the quality of the far-field patterns can be anticipated from the effective number of nonrandom pixels in the input plane. In the remainder of this section we consider potential applications of the method. The method is especially useful in that phase and amplitude are specified independently of each other. Although the examples in the last section demonstrate the approximation of amplitude-only inputs, the mathematical analysis shows that any value of phase (specified as the average value of phase) is permissible. For example, the far-field pattern can be translated by adding a phase slope to the design values of phase values. Therefore, one can use a single phase-only SLM to perform simultaneous and independent beam shaping and beam steering. Two-dimensional scanners can be envisaged that have much more flexibility than current ones. Scanning is not limited to rastered formats, and multiple spots can be formed. For such applications, it should generally not be necessary for

one to use a new set of random numbers each frame. Instead of generating them on the fly, one can store a single frame of random numbers in a video memory. A simple class of functions can also be selected for beam shaping that use a small number of function calculations and memory.

It may be possible to apply this flexibility in scanning to pattern recognition, tracking, and sensing. Currently raster scanning of lasers is used as a way to sense outlines of objects.[27] An edge point is identified by a sudden change in reflectance while scanning. A phase-only SLM-based scanner can locate edges and then adapt the beam shape to enhance the return from the edge. Spots could be contoured and positioned to fit over multiple edges and corners, and this could be used as a feature-based method[28] of tracking or recognition.

The pseudorandom design procedure may have various applications to the design of free-space optical interconnects. Certainly designs can be directly synthesized, and their performance can be readily evaluated; however, their optimality is not guaranteed. We believe that the approach may nonetheless reduce the computation time of other design procedures by providing a better initial guess for the iterative design procedures, especially simulated annealing approaches[4] that start with a pseudorandom array of weights anyway.

What our design procedure does not address is the numerous local minima that one forms by placing the constraint of a phase-only input plane on the design statement. For example, the Dammann grating has a multiplicity of unique solutions consisting of evenly spaced diffraction peaks of equal intensity and phases of 0 or π.[2] One finds the most diffraction-efficient design by exhaustively solving for every possible combination of phase. However, we already know that the most diffracation-efficient designs correspond to those that have a large number of nonrandom pixels. It may be possible for one to use this information to guide such global searches more efficiently.

In optical correlators, pseudorandom phase modulation can represent (typically) real-valued images in the input plane and complex filters in the filter plane with an accuracy that is reasonably modeled. We have proposed a compact phase-only correlator that uses a single phase-only SLM to perform both input- and filter-plane modulation.[29] The pseudorandom encoding is an alternative to nonlinearly transforming real-valued image data to phase. In the filter plane one may use pseudorandom modulation to approximate the full complex matched filter without employing a full complex SLM. Even though the performance will be reduced from that from full complex filters, it may provide useful experimental information for researchers in advance of practical full complex devices.

Pseudorandom phase-only design may well prove useful in many optical processing applications because it is a direct noniterative procedure, and straightforward mathematical expressions are available for the determination of performance bounds of the procedure.

Appendix A

We present a brief sketch of the derivation of Eq. (14), the second-order statistical moment of the far-field intensity $I(f_x)$. Under the assumption of statistically independent pixels, we find that the expectation simplifies to expectation over each pixel when subscripts i, j, k, and l are not equal; however, there are many cases where some or all of the subscripts are equal. In fact, there are 15 possible combinations of equal and nonequal subscripts. These are enumerated in Goodman.[18] The 15 terms can be grouped into seven distinct terms as

$$\begin{aligned}\langle I^2\rangle &= \sum_i\sum_j\sum_k\sum_l \langle A_iA_j^*A_k^*A_l\rangle \\ &= \sum\sum_{i\neq j\neq k\neq l}\sum \overline{A}_i\overline{A}_j^*\overline{A}_k^*\overline{A}_l + \sum_i \langle |A_i|^4\rangle \\ &\quad + 4\,\mathrm{Re}\sum_{i\neq j}\sum \langle |A_i|^2A_i^*\rangle\overline{A}_j \\ &\quad + 2\sum_{i\neq j}\sum \langle |A_i|^2\rangle\langle |A_j|^2\rangle + \sum_{i\neq j}\sum \langle A_i^2\rangle\langle A_j^{*2}\rangle \\ &\quad + 4\sum_{i\neq j\neq k}\sum \langle |A_i|^2\rangle\overline{A}_j\overline{A}_k^* \\ &\quad + 2\,\mathrm{Re}\sum_{i\neq j\neq k}\sum \langle A_i^2\rangle\overline{A}_j^*\overline{A}_k^*. \end{aligned} \tag{A1}$$

Each multiple summation with unequal subscripts can be rearranged as a few summations for which there is no inequality constraint on the subscripts. An example of this is presented in the derivation of Eq. (13) for the double summation that describes the expected value of the far-field intensity. The summations in Eq. (A1) over two, three, and four subscripts are handled in a similar manner, only the algebra is more tedious. One can handle the algebra more easily by using the following tensor shorthand:

$$\begin{aligned} T_{ijkl} &= \sum_i\sum_j\sum_k\sum_l a_ib_jc_kd_l, \\ T_{\substack{ijkl\\ i\neq j\neq k\neq l}} &= \sum\sum_{i\neq j\neq k\neq l}\sum a_ib_jc_kd_l. \end{aligned} \tag{A2}$$

Summations over two and three subscripts are similarly defined as T_{ij} and T_{ijk}. There should be no confusion between the use of the tensor symbol T_{ijkl} and transmittance $T(f_x)$ in this discussion.

The relationships between tensors with subscripts that can be equal and those that cannot be equal are

$$\begin{aligned} T_{\substack{ij\\ i\neq j}} &= T_{ij} - T_{ii}, \\ T_{\substack{ijk\\ i\neq j\neq k}} &= T_{ijk} - T_{iii} - T_{\substack{iij\\ i\neq j}} - T_{\substack{iji\\ i\neq j}} - T_{\substack{ijj\\ i\neq j}}, \\ &= T_{ijk} + 2T_{iii} - T_{iij} - T_{iji} - T_{ijj}. \end{aligned} \tag{A3}$$

The second line of the second equation in Eq. (A3)

follows from the first equation. Similarly, we use Eq. (A3) to derive

$$\sum_{\substack{i\neq j\neq k\neq l}} T_{ijkl} = T_{ijkl} - 6T_{iiii} + T_{iijj} + T_{ijij} + T_{ijji} + 2(T_{iiij} + T_{iiji} + T_{ijii} + T_{ijjj}) - (T_{iijk} + T_{ijik} + T_{ijjk} + T_{ijki} + T_{ijkj} + T_{ijkk}). \quad \text{(A4)}$$

We use Eq. (A4) to represent the first term of Eq. (A1). When $\langle A_i \rangle = a_i = b_i^* = c_i^* = d_i$, as is the case in the first term of Eq. (A1), we can use this symmetry to simplify Eq. (A4) to

$$\sum_{\substack{i\neq j\neq k\neq l}} T_{ijkl} = T_{ijkl} - 6T_{iiii} + 2T_{iijj} + T_{ijji} + 8\,\mathrm{Re}\,T_{iiij} - 4T_{iijk} - 2\,\mathrm{Re}\,T_{ijki}. \quad \text{(A5)}$$

We use Eqs. (A3) and (A5) to evaluate the six terms in Eq. (A1) that have unequal subscripts. Certain terms are recognized as $\langle I \rangle$ and $\langle T \rangle$ [see Eq. (10), (12), and (13)], which then leads to Eq. (14).

The authors are grateful to M. Hudak of Codonics, Inc. and S. Nowell of Alden Electronics, Inc. for providing hard copies of our gray-scale images. This research is sponsored by the Advanced Research Projects Agency and monitored by Rome Laboratory through contract F19628-92-K0021.

References and Notes

1. H. Dammann and K. Gortler, "High-efficiency multiple imaging by means of multiple phase holograms," Opt. Commun. **3,** 312–315 (1971).
2. J. Jahns, M. M. Downs, M. E. Prise, N. Streibel, and S. J. Walker, "Dammann gratings for laser beam shaping," Opt. Eng. **28,** 1267–1275 (1989).
3. F. B. McCormick, "Generation of large spot arrays from a single laser beam by multiple imaging with binary phase gratings," Opt. Eng. **28,** 299–304 (1989).
4. M. P. Dames, R. J. Dowling, P. McKee, and D. Wood, "Efficient optical elements to generate intensity weighted spot arrays: design and fabrication," Appl. Opt. **30,** 2685–2691 (1991).
5. N. C. Gallagher and B. Liu, "Method for computing kinoforms that reduces image reconstruction error," Appl. Opt. **12,** 2328–2335 (1973).
6. R. W. Gerchberg and W. O. Saxton, "Practical algorithm for the determination of phase from image and diffraction plane pictures," Optik (Stuttgart) **35,** 237–250 (1972).
7. If the complex Fourier transform pair is not known, then one FFT would have to be computed. Although such direct methods are much faster than the iterative methods, it is impractical to include electronic calculation of the FFT in real-time optoelectronic processors. Optical Fourier transform and interferometric detection with video cameras can be performed at real-time rates and might be practically employed in the latter case. See, for example, R. W. Cohn, "Adaptive real-time architectures for phase-only correlation," Appl. Opt. **32,** 718–725 (1993).
8. O. Bryngdahl and F. Wyrowski, "Digital holography—computer-generated holograms," in *Progress in Optics,* E. Wolf, ed. (Elsevier, Amsterdam, 1990), Vol. 28, pp. 1–86, and references therein, including reviews by W. J. Dallas and W. -H. Lee.
9. C. B. Burckhardt, "Use of random phase mask for the recording of Fourier transform holograms of data masks," Appl. Opt. **9,** 695–700 (1970).
10. R. Brauer, U. Wojak, F. Wyrowski, and O. Bryngdahl, "Digital diffusers for optical holography," Opt. Lett. **16,** 1427–1429 (1991).
11. R. W. Cohn and R. J. Nonnenkamp, "Statistical moments of the transmittance of phase-only spatial light modulators," in *Miniature and Micro-Optics: Fabrication,* C. Roychoudhuri and W. B. Veldkamp, eds., Proc. Soc. Photo-Opt. Instrum. Eng. **1751,** 289–297 (1992).
12. R. M. Boysel, J. M. Florence, and W. R. Wu, "Deformable mirror light modulators for image processing," in *Optical Information Processing Systems and Architectures,* B. Javidi, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **1151,** 183–194 (1989).
13. J. Amako and T. Sonehara, "Kinoform using an electrically controlled birefringent liquid-crystal spatial light modulator," Appl. Opt. **30,** 4622–4628 (1991).
14. N. Konforti, E. Marom, and S. -T. Wu, "Phase-only modulation with twisted nematic liquid-crystal spatial light modulators," Opt. Lett. **13,** 251–253 (1988).
15. R. W. Cohn, "Random phase errors and pseudorandom phase modulation of deformable mirror spatial light modulators," in *Optical Information Processing Systems and Architectures IV,* B. Javidi, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **1772,** 360–368 (1992).
16. A. Papoulis, *Probability, Random Variables, and Stochastic Processes,* 1st ed. (McGraw-Hill, New York, 1965).
17. The overbar in Eq. (6) is distinguished from the ⟨ ⟩ operator as a way to indicate that the overbarred variable is a result of ensemble averaging. At times, when the meaning is clear, we will interchange the use of the overbar and the ⟨ ⟩ operator in order to improve the readability.
18. J. W. Goodman, *Statistical Optics* (Wiley, New York, 1985), App. C.
19. J. C. Dainty, ed., *Laser Speckle and Related Phenomena,* 2nd ed. (Springer, Berlin, 1984); Ref. 18, Chap. 8.
20. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge U. Press, Cambridge, England, 1986).
21. Equation (24) is identical for both Gaussian and uniform distributions.
22. B. V. K. V. Kumar and L. G. Hassebrook, "Performance measures for correlation filters," Appl. Opt. **29,** 2997–3006 (1990).
23. J. L. Horner, "Metrics for assessing pattern-recognition performance," Appl. Opt. **31,** 165–166 (1992).
24. R. W. Cohn and J. L. Horner, "Effects of systematic phase errors on phase-only correlation," Appl. Opt. (to be published).
25. A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing* (Prentice-Hall Englewood Cliffs, N.J., 1975), Sec. 5.5.
26. F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," Proc. IEEE **66,** 51–83 (1978).
27. H. Kamoda, H. Goto, and K. Imanaka, "Two dimensional optical position sensor using a dual axis miniature optical scanner," in *Miniature and Micro-Optics: Fabrication,* C. Roychoudhuri and W. B. Veldkamp, eds., Proc. Soc. Photo-Opt. Instrum. Eng. **1751,** 280–288 (1992).
28. R. C. Gonzalez and P. Wintz, *Digital Image Processing,* (Addison-Wesley, Reading, Mass., 1977), Sec. 6.5.
29. R. W. Cohn, "Adaptive real-time architectures for phase-only correlation," Appl. Opt. **32,** 718–725 (1993).

US005363186A

United States Patent [19]

Cohn et al.

[11] Patent Number: 5,363,186

[45] Date of Patent: Nov. 8, 1994

[54] **METHOD OF PRODUCING AN OPTICAL WAVE WITH A PREDETERMINED OPTICAL FUNCTION**

[75] Inventors: **Robert W. Cohn; Minhua Liang**, both of Louisville, Ky.

[73] Assignee: **The United States of America as represented by the Secretary of the Air Force**, Washington, D.C.

[21] Appl. No.: **192,475**

[22] Filed: **Jan. 27, 1994**

[51] **Int. Cl.**[5] **G01C 3/08**

[52] **U.S. Cl.** **356/4**; 356/28

[58] **Field of Search** 356/4, 28

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,995,102	2/1991	Ichinose et al.	342/158
5,012,253	4/1991	Schuster et al.	342/203
5,142,289	8/1992	Petersson	342/158
5,184,218	2/1993	Gerdes	358/133
5,187,484	2/1993	Stove	342/200
5,252,981	10/1993	Grein et al.	342/200
5,258,996	11/1993	Fraser et al.	375/1
5,276,636	1/1994	Cohn	364/822

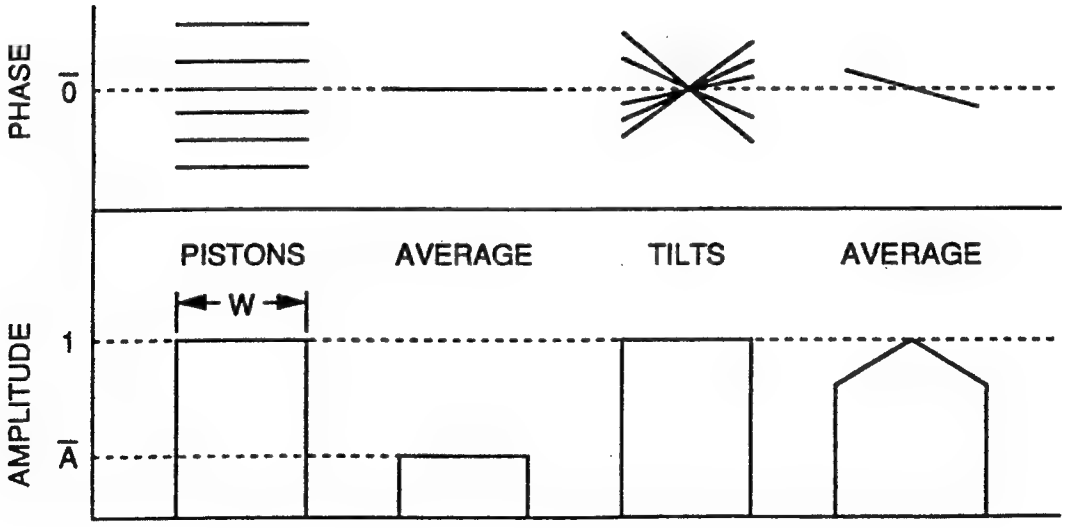
Primary Examiner—Mark Hellner

Attorney, Agent, or Firm—William G. Auton; Jacob N. Erlich

[57] **ABSTRACT**

A five step method is used to produce an optical wave. The first step is to pick the desired far field pattern of the diffractive optical wave. The second step entails transforming the desired far field pattern to the desired source using a fast fourier transform (FFT). In the third step we use the formula $\exp[j\phi_i(\sigma_i)] = \bar{A}_i \exp(j\bar{\phi}_i)$ to find the standard deviation σ_i for each pixel i that produces the desired/average amplitude. Next, we use a pseudo-random number generator to select phases ϕ_i from distributions of standard deviation σ_i for each pixel. Finally, we check the solution with the FFT and our analyses of performance.

9 Claims, 3 Drawing Sheets



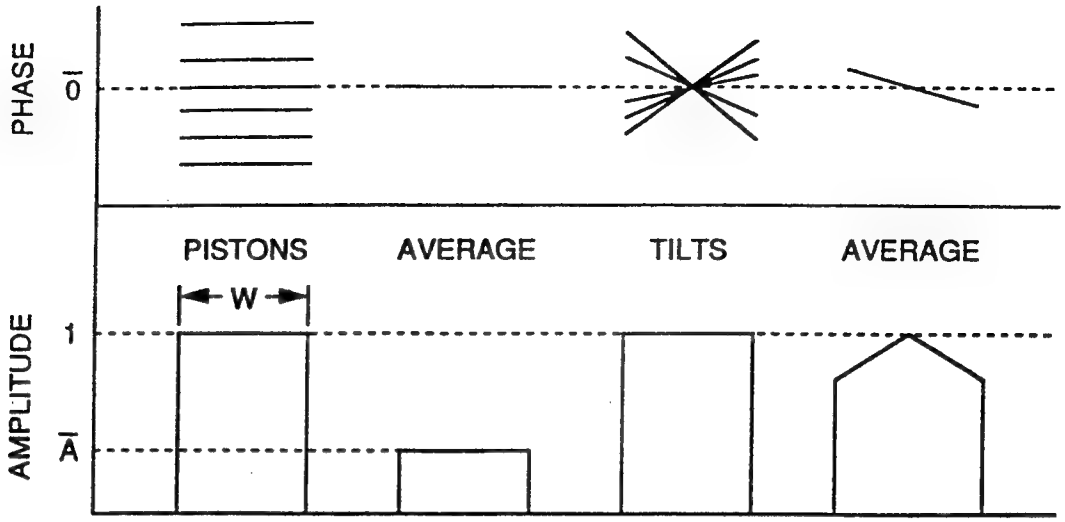

PHASE
0
PISTONS
AVERAGE
TILTS
AVERAGE
W
AMPLITUDE
1
A

FIG. 1

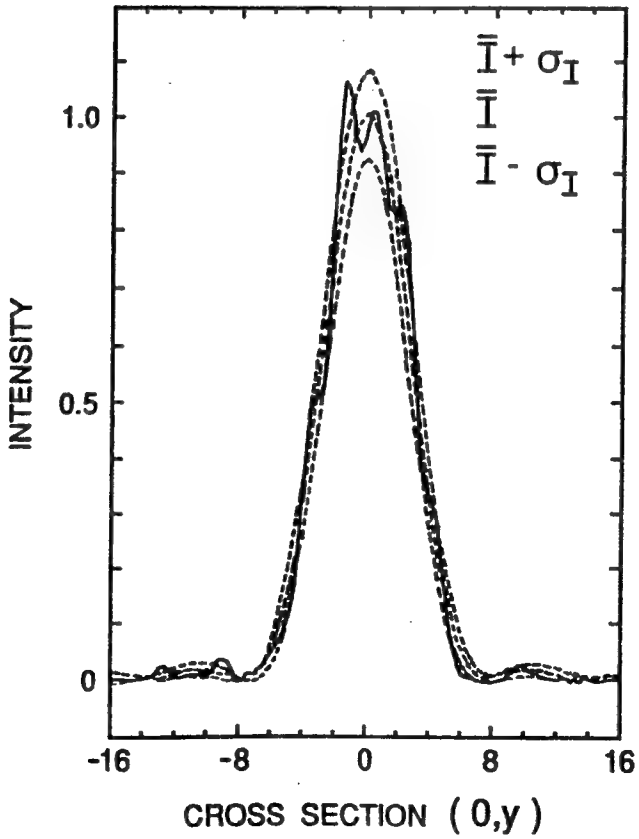

$\bar{I} + \sigma_I$
$\bar{I}$
$\bar{I} - \sigma_I$
1.0
0.5
0
INTENSITY
-16
-8
0
8
16
CROSS SECTION (0,y)

FIG. 2a

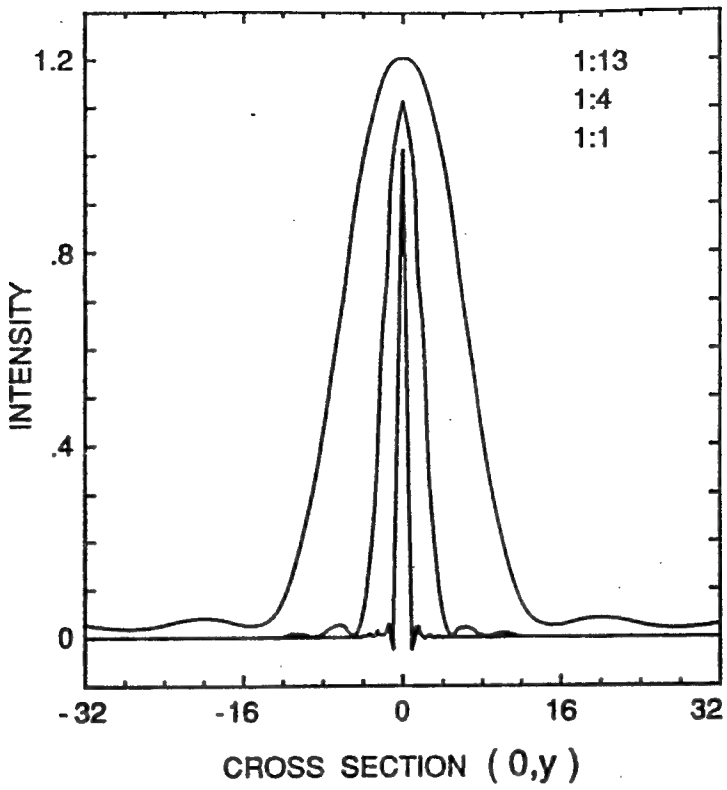

1.2
.8
.4
0
INTENSITY
1:13
1:4
1:1
-32
-16
0
16
32
CROSS SECTION (0,y)

FIG. 2b

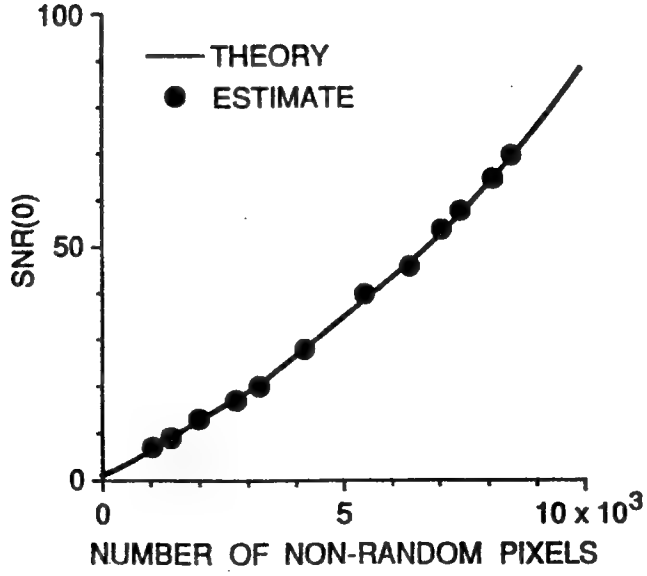

100
50
0
SNR(0)
THEORY
ESTIMATE
0
5
10 x 10^3
NUMBER OF NON-RANDOM PIXELS

FIG. 3

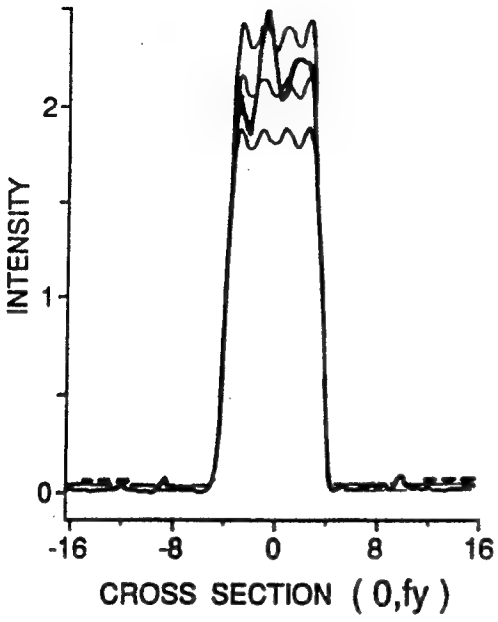


FIG. 4a

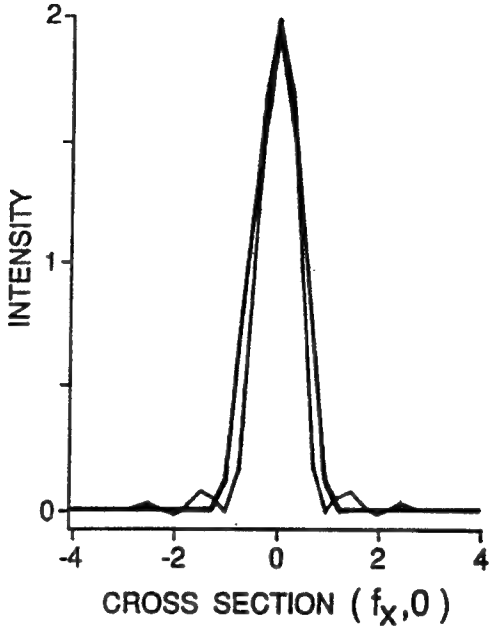


FIG. 4b

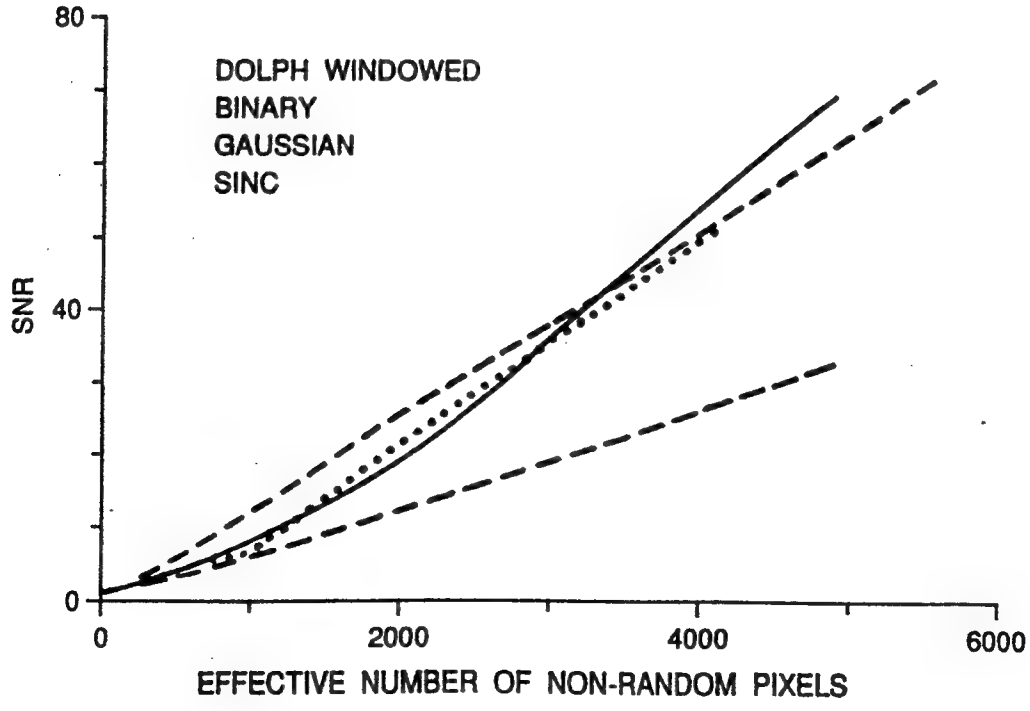


FIG. 5

METHOD OF PRODUCING AN OPTICAL WAVE WITH A PREDETERMINED OPTICAL FUNCTION

STATEMENT OF GOVERNMENT INTEREST

The invention described herein may be manufactured and used by or for the Government for governmental purposes without the payment of any royalty thereon.

BACKGROUND OF THE INVENTION

The present invention relates generally to diffractive optical elements, and more specifically the invention pertains to a method to synthesize phase weights for derived diffraction patterns for optical elements in applications such as a phase-only spatial light modulator (SLM).

Any diffraction pattern can be produced in the Fourier plane by specification of a corresponding input plane transparency. Complex-valued transmittance is generally required but, in practice, phase-only transmittance is used. Many design procedures use numerically intensive, constrained optimization. What is needed instead is to introduce a non-iterative procedure that directly translates the desired, but unavailable, complex transparency into an appropriate phase transparency such that at each pixel the value of phase is pseudorandomly selected from a random distribution whose standard deviation is specified by the desired amplitude, and to derive statistical expressions and use them to evaluate the approximation errors between the desired and achieved diffraction patterns.

This invention is motivated by a desire to design phase-only filters and diffractive elements with a small amount of electronic computation and thereby permit programming of arbitrary spatial modulation at real-time rates. Popular design procedures (e.g. the Dammann grating, simulated annealing, iterative constrained optimization, and other iterative procedures) are only practical if performed off-line due both to the numerical cost of performing Fourier transforms repeatedly and the further cost of evaluating the sensitivity of the transform with respect to a large number of pixels (frequently every pixel of the input plane spatial light modulator.) Of course, the solutions can be precomputed and stored in memory, but only if the number of designs required is not too great.

There are many procedures in the area of computer generated holography, esp. kinoforms that permit direct synthesis of the input plane. These presuppose that the Fourier transform pair between the fully complex-valued input and Fourier planes are known and work by encoding the desired complex values to appropriate phase settings. The direct synthesis design procedures thus allow programming at real-time rates, if the desired Fourier plane pattern is known. The amount of memory is also minimized if the complex valued Fourier transform pair can be written as an easily computed function.

The task producing an optical wave with a predetermined function is alleviated, to some extent, by the systems disclosed in the following U.S. Patents, the disclosures of which are incorporated herein by reference:

U.S. Pat. No. 5,258,996 issued to Fraser, et al;
U.S. Pat. No. 5,187,484 issued to Stove;
U.S. Pat. No. 5,184,218 issued to Gerdes;
U.S. Pat. No. 5,142,289 issued to Peterson;
U.S. Pat. No. 5,252,981 issued to Grein, et al;

U.S. Pat. No. 5,012,253 issued to Schuster;
U.S. Pat. No. 4,995,102 issued to Ichinose, et al; and
U.S. Pat. No. 5,276,636 issued to Cohn, et al.

The patent to Ichinose et al discloses a reversed spiral scanning method used by laser radar. The remaining patents are of interest.

Most frequently the direct procedures operate on a small number of adjacent pixels together as a group that approximates several discrete settings over the complex plane (i.e., cell-oriented encoding). This however reduces the space bandwidth, which is already quite small (say 128×128 pixels) for current spatial light modulators, as compared to traditional fixed pattern holographic and diffractive optical elements. The procedure of the present invention is also a direct method, but one for which a continuous value of phase is selected for each individual pixel independent of all other settings (i.e., point-oriented encoding.)

SUMMARY OF THE INVENTION

The present invention includes a five step method to synthesize the pixel phase and amplitudes of an optical wave. The first step is to pick the desired far field pattern of the diffractive optical wave. The second step entails transforming the desired far field pattern to the desired source using a fast fourier transform (FFT). In the third step we use the formula $\overline{\exp[j\phi_i(\sigma_i)]} = \bar{A}_i \exp(-j\bar{\phi}_i)$ to find the standard deviation σ_i for each pixel i that produces the desired/average amplitude. Next, we use a pseudo-random number generator to select phases ϕ_i from distributions of standard deviation σ_i for each pixel. Finally, we check the solution with the FFT and our analyses of performance. As mentioned above, phase-only holograms are often designed to produce far-field diffraction patterns that approximate desired patterns. Iterative search for the optimal transform pair under the constraint of phase-only modulation is not always required. In the present invention, we instead select a desired far-field pattern which specifies the source distribution. Amplitude and phase are encoded by controlling the randomness of phase across the source.

The method is analogous to placing a variable roughness diffuser over the source. Increasing roughness at a pixel decreases its coherent contribution to the far-field pattern.

Any array of random phase pixels has an average far-field pattern that is identical to the Fourier transform of the average pixel transmittances. It follows that the average far-field pattern is approximated by setting each pixel to a value randomly selected from its individual phase ensemble.

This process described above can be used by either a radar or an optical phased array steering system, and applied to a feature based tracking system for automated production line technology.

It is an object of the present invention to produce a method to design phase-only diffractive optical elements.

It is another object of the present invention to synthesize desired phase weights for desired diffractive patterns.

These objects together with other objects, features and advantages of the invention will become more readily apparent from the following detailed description when taken in conjunction with the accompanying drawings wherein like elements are given like reference numerals throughout.

DESCRIPTION OF THE DRAWINGS

FIG. 1 is an illustration of expected pixel and SLM transmittance of the phase and amplitudes of random phase-only pixels;

FIG. 2*a* is a chart of the diffraction pattern of a phase-only elliptical aperture; (cross section along long axis);

FIG. 2*b* is a chart of the performance of a 128 by 128 SLM for various elliplicities;

FIG. 3 is a chart of the signal to noise ratio at diffraction peak as a function of the number of non-random pixels for pseudorandom encoding of binary amplitudes;

FIG. 4 is a set of two charts (a) and (b) of the diffraction pattern of a phase-only approximated Dolph-windowed since anodization where (a) cross section is across vertical axis (O,f_y) and the dashed horizontal lines indicate the saturation (full-white) level; and

FIG. 5 is a chart of the SNR as a function of effective number of non-random pixels for pseudorandom encoding of continuous amplitudes, were the SNR(O) is plotted for gaussian and binary, and average SNR across passband is plotted for Dolph and sinc designs, and the sinc is a rectangularly symmetric function and the gaussian is a circularly symmetric function.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

The present invention includes a process for designing phase-only diffractive optical elements and to synthesize phase weights for desired diffraction patterns.

The present method directly follows from mathematical models (presented below) of far-field diffraction from arrays of randomly-phased point sources. More specifically, the phases are treated as independent and non-identically distributed random variables. It is convenient for analysis and the design procedure to represent the random variables with a single probability density function that is varied in its mean and standard deviation. Our analyses shows that on average, the pixels with phases drawn from large standard deviation distributions behave as if they have a smaller amplitude transmittance. This result was noted in an average contrast loss for arrays of phase-only pixels that have identically distributed phase errors. This result is a special case of that for non-identically distributed phases.

The ideal SLM for our procedure modulates phase continuously over a full 2π range. Regularly spaced pixels are assumed in the examples below but are not required in our theory. These characteristics are typical of those anticipated for the piston-modulating (or flexure-beam) deformable mirror device and is achievable with birefringent and twisted nematic liquid crystals. For binary and multi-level optics the quantized values of phase can be modeled as statistical departures from the desired analog phase and their additional effect on the diffraction pattern can be estimated.

The design procedure of the present invention specifies the degree of randomness (i.e. standard deviation) of phase at each pixel in order to approximate arbitrary values of amplitude. This is analogous to placing a diffuser of spatially varying roughness over the input plane. Increasing roughness at a pixel decreases its coherent contribution to the far-field pattern. The remaining incoherent, or diffused light is spread over the entire diffraction pattern and contributes a noise background that is frequently referred to as speckle. The design phase at any pixel is specified as the expected value of its random phase distribution. Thus the full complex input plane can be viewed as a cascade of a deterministic phase screen with a variable roughness phase screen.

This interpretation follows from the statistical expectation of a complex exponential of a random argument which is schematically illustrated in FIG. 1 and derived in the following section. FIG. 1 shows the spatial phase and amplitude for two types of phase-only pixels:

1. piston-only, those that produce a single value of phase across the pixel width; and
2. tilt-only, those that produce a linearly-varying value of phase across the pixel width.

The design procedure is concerned with the former structure, and the latter is provided as a further example of the concept. The leftmost of the four plots of amplitude and phase represents a random phase pixel. Because phase is a random variable we show an ensemble of phase values. The ensemble average of the random complex exponential produces the result in the second column—an average value of phase and a loss in amplitude transmittance. The example of pixels with random tilts shows a similar result in that the average amplitude transmittance decreases with increasing phase fluctuations that directly correspond with distance from the pivot point.

Because both expectation and Fourier transform are linear operations then the expected complex amplitude of the far-field pattern of the piston-only pixel, as well as an array of pixels, is the Fourier transform of the expected transmittance. The expected intensity of the far-field pattern consists of the magnitude squared of the complex amplitude plus a broad pedestal due to the average intensity of speckle.

In our design procedure, rather than average many trials, we instead select a single value of phase from the ensemble for each pixel. The expected complex amplitude and intensity of the far-field pattern for this situation is mathematically equivalent to that described in the previous paragraph. More important, the actual far-field pattern approximately resembles the average pattern that results from the coherent summation of a large number of random wavefronts. Thus, the formation of the far-field pattern can be viewed as a physical example of the central limit theorem, i.e. the so called "law of large numbers".

The following definitions for arrays of piston-only pixels are used in the development of the statistical expressions. The complex transmittance of an individual pixel located at position X_i will be written as $a_i(X)$ and the transmittance of the array of the N individual pixels is

$$t(x) = \sum_{i=1}^{N} a_i(x) = \sum_i r(x - x_i)\exp(j\psi_i) \tag{1}$$

where Ψ is the phase shift produced by the i'th pixel of the SLM. The abbreviation r(X)=rect(x/w) has been used where w is the width of each pixel. Any inactive area between the pixels has been treated as non-reflecting and the amplitude and phase of each pixel have been defined in local coordinates centered around x=O and then shifted to pixel locations x_i. The intensity of the far-field diffraction pattern is written

$$I(f_x)=T(f_x)T^*(f_x)=\mathfrak{F}\{t(x)\oplus t(x)\} \tag{2}$$

where $T(f_x)$ is the Fourier transform of the transmittance t(x) and where $\oplus$ indicates the correlation integral

$$a(x)\oplus b(x)=\int a(x'+x)b^*(x')dx' \tag{3}$$

The squared intensity spectrum expressed in terms of the fourth-order autocorrelation of SLM transmittance is

$$I^2=TT^*T=\mathfrak{F}\{[t(x)\oplus t(x)]\oplus[t(x)\oplus t(x)]\} \tag{4}$$

Unlike convolution, the order in which correlations are performed affects the result, and thus the brackets are required in eq. 4.

The expected value of a complex phasor of a random argument is frequently referred to as the characteristic function of the random variable Ψ

$$M'(\omega)= \tag{5}$$

$$<\exp(j\omega\psi)> = \int_{-\infty}^{\infty} \exp(j\omega\psi)\rho_\psi(\psi)d\psi = 2\pi\mathfrak{F}^{-1}\{\rho_\psi(\psi)\}$$

where $<>$ is our symbol for the ensemble average (i.e. expectation) operator and $\mathfrak{F}$ is the Fourier transform operator, and $\rho\Psi$ is the probability density function of the random variable Ψ. If the phase Ψ_i are independent random variables then the expected value of pixel transmittance $a_i(x)$ is

$$\begin{aligned}\overline{a_i}(x) &= <a_i(x)> = <r(x-x_i)\exp(j\psi_i)> \\ &= r(x-x_i)M_i'(1) \\ &= r(x-x_i)\exp(j\psi_i)M_i(1)\end{aligned} \tag{6}$$

where M'_i is the characteristic function of Ψ_i and M_i is the characteristic function of the unbiased values of phase $\Psi_i - <\Psi_i>$. The last line of eq. 6 has been written to identify the magnitude and phase components and corresponds with the second average piston case in FIG. 1.

We considered two specific probability distributions for phase Ψ_i. For gaussian distributed phase of standard deviation σ_i

$$M_i(1) = \exp\left(-\frac{1}{2}\sigma_i^2\right) \tag{7}$$

and for uniformly distributed phase with total spread

$$v_i = \sqrt{12}\ \sigma_i \tag{8}$$

then

-continued

$$M_i(1) = \operatorname{sinc}\left[\frac{v_i}{2\pi}\right] \tag{9}$$

For either distribution the average amplitude transmittance at each pixel i can be controlled explicitly by selection of the value of the standard deviation σ_i. In digital simulations we prefer the uniform distribution since most gaussian random number generators are derived by performing additional numerical operations on uniform samples. A further advantage of using uniform statistics is that the total spread never need exceed 2π, whereas with gaussian statistics the standard deviation can be infinite, and which can lead to overflow and underflow errors if not handled carefully.

The expected value of eq. 1, using either probability distribution, is the expected SLM transmittance

$$\overline{t}(x) = \sum_{i=1}^{N} <a_i(x)> = \sum_i r(x-x_i)p_i^{\frac{1}{2}}\exp(j\psi_i) \tag{10}$$

where the parameter

$$P_i=M_i^2(1) \tag{11}$$

is used to simplify subsequent expressions. Note that because the expectation and Fourier transform operators are both linear the expected far-field transmittance (more precisely, the angular spectrum) is

$$\overline{T}(f_x)=\mathfrak{F}\{\overline{t}(x)\} \tag{12}$$

The nomenclature in eq. 12 of using lower-case letters for space domain variables and upper-case variables for frequency domain variables will be used throughout this discussion.

The most general expectation for the intensity of the far-field diffraction pattern SLMs with statistically independent pixels follows from the first equality in eq. 1 and eq. 2. It is

$$\begin{aligned}\overline{I}(f_x) &= \sum_i\sum_j <A_i(f_x)A_j^*(f_x)> \\ &= \sum_i\sum_j <A_i><A_j^*> + \sum_i <|A_i|^2> \\ &= \sum_i\sum_j <A_i><A_j^*> - \sum_i |<A_i>|^2 + \sum_i <|A_i|^2> \\ &= |\overline{T}|^2 + \sum_i [<|A_i|^2> - |\overline{A_i}|^2]\end{aligned} \tag{13}$$

where the intermediate results indicate how independence was used to simplify the expression. This expression shows that the expectation nearly separates into $|<T>|^2$ except for terms involving the autocorrelation of the input plane pixels (i.e. terms for which i=j.) Under the same assumption of independent pixels the most general expression for the squared far-field intensity is

$$\begin{aligned}<I^2> &= \sum_i\sum_j\sum_k\sum_l <A_iA_j^*A_k^*A_l> \\ &= 2[\overline{I}^2 - |\overline{T}|^4] + |\overline{T}^2 + \sum_i (<A_i^2> - \overline{A_i}^2)|^2 + \\ &\quad 4Re\left[\overline{T}^*\sum_i (<|A_i|^2A_i> - <A_i^2>\overline{A_i}^* + 2|\overline{A_i}|^2\overline{A_i} - 2<|A_i|^2>\overline{A_i})\right] +\end{aligned} \tag{14}$$

-continued

$$\sum_i [<|A_i|^4> - 6|\overline{A_i}|^4 + 8<|A_i|^2>|\overline{A_i}|^2 - |<A_i^2>|^2 - 2<|A_i|^2>^2 + 4Re(<A_i^2>\overline{A_i}^{*2} - <|A_i|^2A_i>\overline{A_i}^*)]$$

It was found by substituting eq. 1 into eq. 4 and then taking the expectation of eq. 4. The second equality was found by a procedure similar to that shown in eq. 13. The standard deviation of intensity $\sigma_1(f_x)$ is then directly found using the well known result

$$\sigma_I^2 = <I^2> - \overline{I}^2 \tag{15}$$

for the specific case of piston-only, non-identically distributed random phase eqs. 13–15 simplify to

$$I(f_x) = \mathfrak{F}\left\{ \overline{r}(x) \oplus \overline{r}(x) + [r(x) \oplus r(x)] \sum_i q_i \right\} \tag{16}$$

$$= |\overline{T}(f_x)|^2 + R^2(f_x) \sum_i q_i$$

where

$$q_i = 1 - p_i \tag{17}$$

and

$$\sigma^2(f_x) = \overline{I^2} - 2|\overline{T}|^4 + |\overline{T^2} - T_A|^2 - 4Re[\overline{T}T_B^*] - \sigma_A \tag{18}$$

We introduce a shorthand for defining repeated autocorrelations of the rect function r(x)

$$g_n(x) = [r(x) \oplus \ldots \oplus r(x)]_n \tag{19}$$

where n indicates the number of rect functions (i.e. n−1 integrals). Eq. 18 is completely specified with the additional definitions for the gaussian distribution

$$t_A(x) = \sum_i g_2(x - 2x_i) \exp(j2\psi_i)\, q_i p_i \tag{20}$$

$$t_B(x) = \sum_i g_3(x - x_i) \exp(j\psi_i)\, q_i^2 p_i^2$$

$$g_A(x) = g_4(x) \sum_i q_i^4$$

and for the uniform distribution

$$t_A(x) = \sum_i g_2(x - 2x_i) \exp(j2\overline{\psi}_i)(p_i - d_i) \tag{21}$$

$$t_B(x) = \sum_i g_3(x - x_i) \exp(j\overline{\psi}_i)(q_i - p_i + d_i) p_i^2$$

$$g_A(x) = g_4(x) \sum_i (q_i - 3p_i + 6p_i^2 - 4p_i d_i + d_i^2)$$

where

$$d_i = \mathrm{sinc}\left[\frac{v_i}{\pi}\right] \tag{22}$$

The expected intensity pattern (eq. 16) contains the desired design intensity pattern (the magnitude square of eq. 12) plus an additional term [proportional to $\sigma_2(f_x) = R^2(f_x)$] that we associate with the average level of speckle intensity (and which is often referred to as "halo" or "pedestal".) This second term is proportional to the far-field diffraction pattern intensity of a single pixel. Examination of eq. 18 also reveals that each of its terms has the common factor $G4(f_x) = G_2^2(f_x)$. Thus for a standard definition of single-to-noise ratio

$$SNR(f_x) = \frac{\overline{I}(f_x)}{\sigma_I(f_x)} \tag{23}$$

the term $G_2(f_x)$ cancels out and is independent of spatial frequency. The reciprocal of eq. 23 basically describes the relative approximation error (especially when σ_I is small compared to I and the contribution of pedestal can be partially ignored.)

Another type of signal-to-noise ratio, that is quite common in describing the quality of an optical correlation peak, is the ratio of peak intensity to background noise level, and is often referred to as peak-to-noise ratio (PNR). The analysis above permits calculation of this as well, however, unless the pixels are point sources the pixel elements factor $G2(f_x)$ needs to be considered and so there will be different values of signal-to-noise-ratio depending on the pixel fill factor. The steps of the design procedure of the present invention are:

1) Specify the desired far-field pattern T(f).
2) Fourier transform to the desired source distribution t(x) such that for each pixel i: invert eq. 9 (preferred, or eq. 7) to find σ_i.
3) Select a pseudorandom number, appropriately scaled by σ_i to represent the random phase ϕ_i. Fourier transform the array of random phase-only pixels to evaluate the actual far-field pattern. If desired, compare the actual far-field $I(f_x)$ with eqs. 16, 18 and 23.

Several designs have been simulated and analyzed, two of which are presented in some detail. In every case the design is for a 128×128 pixel phase-only SLM. It is represented as a 128×128 array of samples embedded in the center of a 512×512 array of zeroes. A fast Fourier transform (FFT) routine is used to approximate the Fourier transform of the array and of the individual terms comprising the expectations in eqs. 16 and 18.

The uniform random number generator DRNUNF (from the IMSL numerical software library) with initialization subroutine RNOPT(6) was used. We mention this because initially there were problems using the random number generator RAN1 Specifically, when we attempted to verify eqs. 16 and 18 by Monte Carlo estimation (that is ensemble averaging of a large number of identical designs generated with different random number seeds) the estimates sometimes converged to different results than the equations. The discrepancies were especially obvious at harmonically related sets of spatial frequencies. This may indicate that RAN1 produced correlated sequences in these experiments. These problems however were not discernable in individual design runs because of the magnitude of the random fluctuations.

When using DRNUNF we did find that plots of Monte Carlo estimates of the expected value and standard deviation of intensity, except for small fluctuations, appear to closely match our expressions when 1000 design runs are used in the estimate. We also checked that the relative error between the theory and estimate decreases with increasing number of design runs and that this improvement is roughly proportional

to the square root of the number of design runs used in the estimate, as would be expected. Our closest result, using 10^5 design runs, converged to the theoretical expressions with fluctuation of less than 0.4% standard deviation.

A phase modulation was developed to approximate an elliptically shaped aperture of eccentricity 6:1. FIG. 2*a* presents the diffraction pattern along the vertical axis. FIG. 2*b* is a chart of the expected intensity profiles. The 1+01 ratio is nearly equivalent to the diffraction pattern along the horizontal axis for all values of eccentricity. FIGS. 2*a* and 2*b* together are examples of beam shaping by this method. FIG. 2*a* presents quantitative information along the vertical axis. The light curves depict the error bars of ± 1 added to the expected intensity. These provide a rough idea of the magnitude of the actual fluctuations, both of the coherent peak and the speckle dominated sidelobe region.

It should be apparent that fluctuations, and thus approximation errors will be smaller for those binary amplitude designs that use a greater number of non-random pixels. This relationship can be seen in eq. 23 which simplifies to

$$SNR(0) = \frac{N_N^2 + N_R}{\sqrt{2\,N_R N_N^2 - N_R + N_R^2}} \approx \frac{N_N}{\sqrt{2\,N_R}} \tag{24}$$

the zero frequency where the diffraction intensity peaks. N_R is the number of randomly modulated pixels on the SLM and N_N is the number of non-randomly modulated pixels and their sum is the total number of SLM pixels, 16,384. The approximation in eq. 23 is valid except for SNR close to unity; for instance, for 1000 non-random pixels the relative error is roughly 1% and the SNR is 5.7. FIG. 3 plots eq. 24 against 500 run Monte Carlo estimates of SNR(0). These estimates correspond to elliptical aperture designs with a major axis of 128 pixels and eccentricity ranging from 1 to 13 and circular apertures of diameter from 42 to 128 pixels. For reference, there are 2016 non-random pixels in the ellipse of eccentricity 6:1. This data point on FIG. 3 is nearly indistinguishable from the point for a circular aperture of diameter 50 pixels and which contains 1976 pixels.

A second performance measure describing the quality of the far-field diffraction pattern is the peak intensity to noise ratio which we choose to define here as

$$PNR = \frac{\bar{I}(0)}{\sigma_N(f_x)} \approx \frac{N_N^2}{N_R} \approx 2\,SNR^2(0) \approx \frac{\bar{I}(0)}{\bar{I}(f_x)} \tag{25}$$

where f_x is assumed to be a frequency in the sidelobe region. The approximations follow by assuming the expected transmittance is negligible (i.e. $\langle T(f_x)\rangle = 0$ in eqs. 16 and 18) with respect to the speckle intensity in the sidelobe region. Additionally, we set the ratio of $G_2(0)/G_2(f_x)$ to unity, mainly to more dramatically indicate the simple relationship between PNR and SNR(0). In our simulations, for which we choose f_x as the Nyquist frequency [i.e. the point half way between the (0,0) and the (1,1) diffraction order], the first approximation indicates that the average level of speckle in the sidelobe regions is equal to the standard deviation of intensity. This result is not unexpected based on previous observations that the intensity of fully developed speckle patterns is exponentially distributed and that the standard deviation of this distribution is equal to its mean.

We have also evaluated the design procedure for the approximation of an apodized input plane. Two dimensional gaussian, airy pattern and sinc functions with varying aspect ratioes have been synthesized. The sinc and airy functions were chosen to produce brickwall and top-hat diffraction patterns. However, the finite spatial extent of the SLM severely truncates the ideal functions and produces significant in-band ripple (Gibbs phenomena) and large sidelobes. The only feature to recommend this design is the extremely rapid transition band.

The transition bandwidth can be readily traded off with ripple and sidelobe level by the window design method that is widely used in the design of finite impulse response digital filters. In this method truncation effects are controlled by multiplying the ideal, infinite extent functions by an amplitude tapered window in place of a rect function. One of the best windows for designing uniform intensity patterns is the Dolph-Chebyshev function. It is considered optimal in the sense that its Fourier transform produces the minimum mainlobe width for a given sidelobe level. The parameter α specifies the sidelobe level of 20α in dB.

We have used the window method to design a pseudorandom phase modulation that will diffract into a close approximation to a brickwall shape. The design function is a one dimensional sinc function multiplied by a Dolph window in both coordinates. Our general design goal for this example is to produce a diffraction with the largest aspect ratio (or eccentricity) possible that still reasonably resembles a brickwall shape. We feel that for a 128×128 SLM this goal is met by the function sinc(4x) multiplied by Dolphs in x and y that both have α equal to 1.3.

This result is presented in FIG. 4.

FIG. 4*a* makes clear that increasing further also increases the intensity of speckle and overall passband ripple is not reduced because the random fluctuation is already larger than the ripple for the Dolph function. The aspect ratio of this diffraction pattern is 6.7:1 at the half power as measured from the plots in FIG. 4. FIG. 4 also shows that Dolph windowing in the vertical direction has widened the diffraction pattern by about 25% at the half power points.

In order to compare the performance of various design functions with the phase-only encoding procedure we introduce the definition of effective number of random pixels

$$\hat{N}_R = \sum_i^N q_i \tag{26}$$

Eq. 27 is identically the amplitude of the noise pedestal term in eq. 16 and it is identical to N_R for the case of binary modulation. FIG. 5 compares SNR for the various functions studied. The curves demonstrate a rough correspondence, with the apodized functions being more closely clustered than the binary amplitude functions. Also not shown is the circularly symmetric sapproximation to an airy pattern on the input. This curve very nearly overlays the sinc curve. For reference, the SNR of the design in FIG. 4 has SNR of roughlt 8 which corresponds to an effective number of 1000 non-random pixels.

The present invention is a method, based on properties of random phase having spatially varying statistics,

that approximates fully-complex input plane modulation. Diffraction patterns from pseudorandom phase-only modulation can be as energy efficient as any passive, fully complex modulation; with performance loss arising from the addition of a nearly-uniform level speckle background. Designers can use the method to specify diffractive optical elements directly from Fourier transform relationships between the input plane and far-field. The performance of any design can be readily evaluated and the quality of the far-field patterns can be anticipated from the effective number of non-random pixels in the input plane.

The method is especially useful in that phase and amplitude are specified independently of each other. Although the examples in the last section demonstrate approximation of amplitude-only inputs, the mathematical analysis shows that any value of phase (specified as the average value of phase) is permissible. For example, the far-field pattern can be translated by adding a phase slope to the design values of phase values. Therefore, a single phase-only SLM can be used to perform simultaneous and independent beam shaping and beam steering. Two dimensional scanners can be envisioned that have much more flexibility than previously. Scanning is not limited to rastered formats and multiple spots can be formed. For such applications, it should generally not be necessary to use a new set of random numbers each frame. Instead of generating them on-the-fly, a single frame of random numbers can be stored in a video memory. A simple class of functions can also be selected for beam shaping that use a small number of function calculations and memory.

It may be possible to apply this flexibility in scanning to pattern recognition, tracking and sensing. Currently raster scanning of lasers is used to sense outlines of objects. An edge point is identified by a sudden change in reflectance while scanning. A phase-only SLM based scanner can locate edges and then adapt the beam shape to enhance the return from the edge. Spots could be contoured and positioned to fit over multiple edges and corners and this could be used as "feature based" method of tracking or recognition.

The pseudorandom design procedure may have various applications to the design of free space optical interconnects. Certainly designs can be directly synthesized, and their performance can be readily evaluated; however, their optimality is not guaranteed. We feel that the approach may nonetheless reduce the computation time of other design procedures by providing a better initial guess for the iterative design procedures; especially simulated annealing approaches that start with a pseudo-random array of weights anyway.

In optical correlators, pseudo-random phase modulation can represent (typically) real-valued images in the input plane and complex filters in the filter plane with an accuracy that is reasonably modeled. We have proposed a compact phase-only correlator that uses a single phase-only SLM to perform both input and filter plane modulation. The pseudorandom encoding is an alternative to the nonlinearly transforming real-valued image data to phase. In the filter plane pseudorandom modulation may be used to approximate the full-complex matched filter without employing a full-complex SLM. Even though the performance will be reduced from that from full complex filters, it may provide useful experimental information for researchers in advance of practical full-complex devices.

Pseudorandom phase-only design may well prove useful in many optical processing applications because it is a direct non-iterative procedure, and straightforward mathematical expressions are available for determining performance bounds of the procedure.

While the invention has been described in its presently preferred embodiment it is understood that the words which have been used are words of description rather than words of limitation and that changes within the purview of the appended claims may be made without departing from the scope and spirit of the invention in its broader aspects.

What is claimed is:

1. A method of producing an optical wave having a predetermined optical function and wherein said method comprises the following steps:

selecting a desired far-field pattern for said optical wave containing pixel phases and pixel amplitudes that represent said predetermined optical function;

performing a fast Fourier transform on said desired far-field pattern to yield, a desired source distribution description for said optical wave;

determining a standard deviation σ_i for each pixel i that produces the predetermined optical function to yield thereby the pixel amplitudes Ai; and

generating the pixel phases ϕ_i from random distributions of the standard deviation σ_i for each pixel i.

2. A method as defined in claim 1, wherein said determining step comprises using a formula given by

$$\overline{\exp[j\phi_i(\sigma_i)]}=\bar{A}_i\exp(j\bar{\phi}_i)$$

to find the standard deviation σ_i for each pixel i that produces a desired average amplitude.

3. A method as described in claim 1, wherein said generating step comprises using a pseudo-random number generator to select phases ϕ_i from distributions of standard deviation σ_i for each pixel.

4. A method as described in claim 2, wherein said generating step comprises using a pseudo-random number generator to select phases ϕ_i from distributions of standard deviation σ_i for each pixel.

5. A method as described in claim 2, wherein said generating step comprises using a pseudo-random number generator to select phases ϕ_i from distributions of standard deviation σ_i for each pixel.

6. A method as defined in claim 1, in which after using the selecting, performing, determining and generating steps to produce said optical wave with an actual configuration, further including the steps of:

comparing the actual configuration of the optical wave produced by the method with the predetermined optical function to evaluate the method.

7. A method as defined in claim 2, in which after using the selecting, performing, determining and generating steps to produce said optical wave with an actual configuration, further including the steps of:

comparing the actual configuration of the optical wave produced by the method with the predetermined optical function to evaluate the method.

8. A method as defined in claim 3, in which after using the selecting, performing, determining and generating steps to produce said optical wave with an actual configuration, further including the steps of:

comparing the actual configuration of the optical wave produced by the method with the predetermined optical function to evaluate the method.

9. A method as defined in claim 4, in which after using the selecting, performing, determining and generating steps to produce said optical wave with an actual configuration, further including the steps of:

comparing the actual configuration of the optical wave produced by the method with the predetermined optical function to evaluate the method.

* * * * *

Pseudorandom phase-only encoding of real-time spatial light modulators

Robert W. Cohn and Minhua Liang

We previously proposed a method of mapping full-complex spatial modulations into phase-only modulations. The Fourier transform of the encoded modulations approximates that of the original complex modulations. The amplitude of each pixel is encoded by the property that the amplitude of a random-phasor sum is reduced corresponding to its standard deviation. Pseudorandom encoding is designed for phase-only spatial light modulators that produce 360° phase shifts. Since such devices are rare, experiments are performed with a 326° modulator composed of two In Focus model TVT6000 liquid-crystal displays. Qualitative agreement with theory is achieved despite several nonideal properties of the modulator.

Key words: Optical information processing, spatial light modulators, liquid-crystal televisions, phase-only filters, laser speckle, rough-surface scattering, statistical optics, binary and diffractive optics, phased arrays, measurements of phase.

1. Introduction

Pseudorandom phase-only encoding[1] is a method for designing phase-only filters and diffractive optical elements that approximately produces the same Fraunhofer diffraction pattern as would result from a desired, but unrealizable, full-complex filter. The encoding procedure adds amplitude control to the phase-only filter[2] through the addition of phase offsets that have specified statistical properties. The randomness of the phase offset at a given pixel, as measured by the standard deviation of this random variable, determines the effective attenuation caused by the pixel insofar as it describes its effect on the Fraunhofer pattern. Each pixel can be programmed in this way to have a specified value of phase and effective amplitude. The fact that the resulting diffraction pattern approximates the desired diffraction pattern from the full-complex modulation is a result of the law of large numbers. In this situation the individual Huygens wave fronts from all spatially separated pixels in the modulator plane all coincide and coherently add together across the Fraunhofer plane (as illustrated in Fig. 1). If the amount of randomness present in the input pixels is not too large, then a coherent reconstruction is observed that approximates the diffraction pattern from the full-complex modulation. If the randomness is too great, only a speckle pattern is observed. We also show that the quality of the reconstruction, in terms of how well it approximates the desired diffraction pattern, is closely related to the average intensity transmittance of the desired full-complex modulation. Since this metric is calculated directly from the full-complex modulation, it can be used in advance of performing the encoding to determine whether or not the phase-only modulation provides adequate performance for the given application.

The method is especially well suited for real-time programming of spatial light modulators (SLM's). The mapping, since it requires one function calculation (or table lookup) per pixel, can be computed at the frame rate by a serial electronic processor. This is in contrast to many diffractive-optic and filter-design procedures that focus on optimal synthesis of a desired diffraction pattern under the constraint of phase-only modulation. These latter approaches are appropriate for the design of fixed-pattern filters; however, since they are numerically intensive, they usually cannot be performed in real time. For some applications it may be adequate to precompute the mapping offline, but for other applications (e.g., those in which the number of precomputed images exceeds the amount of available memory or those in which

R. W. Cohn is with the Department of Electrical Engineering, University of Louisville, Louisville, Kentucky 40292. M. Liang is with Neuristics Corporation, Baltimore, Maryland 21204-2316.

Received 9 June 1995; revised manuscript received 30 November 1995.

0003-6935/96/000001-11$06.00/0

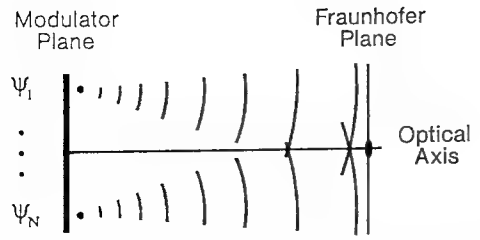


Fig. 1. Fraunhofer diffraction geometry illustrating superposition of a large number of wave fronts from independently phased, equal-intensity point sources.

the desired full-complex function is not necessarily known in advance), it may only be possible to program the SLM by a real-time algorithm such as pseudorandom encoding.

Several devices can in principle produce a 2π phase-modulation range, but far fewer are available in practice. Parallel-aligned nematic liquid crystals (also referred to as birefringent liquid crystals) have been known to produce pure phase modulation for some time.[3] However, such devices have become available only recently. These include devices from Hughes, Hamamatsu, and Meadowlark Optics. Only the optically addressed SLM from Hamamatsu has a 2π modulation range (according to their technical data sheet dated October 1993). Epson researchers have also demonstrated programmable kinoforms using a birefringent liquid-crystal television (LCTV) panel.[4] However, the device is a custom-modified version of a commercially available twisted-nematic display, and this device is not available outside Epson. Using commercially available twisted-nematic LCTV's, several researchers have identified phase-mostly modulation characteristics.[5–7] These are achieved by the proper selection and orientation of polarizers and/or wave plates. Most interesting is the eigenpolarization mode in which an elliptical polarization is transformed into an identical polarization independent of the voltage applied across the liquid crystal.[6] Phase shifts of 195° and better than ±4% uniformity of amplitude were achieved for a panel from the In Focus model TVT6000 video projector. Soutar and Monroe identified a configuration that produces 2.5π of modulation for LCTV panels from an Epson E1020 video projector.[7] However, Soutar and Lu later noted that the modulation characteristics can vary substantially across this device.[8] The (phase-only) deformable mirror device produces phase shifts through the displacement of micromechanical pixels in a direction parallel to the optical axis.[9] However, fully addressable devices with this type of pixel have not yet been successfully demonstrated.

In spite of the limited availability of 2π modulation it is still possible to experimentally demonstrate the method of pseudorandom encoding on current modulators and to produce diffraction patterns that are quite similar to those predicted for 2π modulators. For the TVT6000 LCTV, several nonideal properties must be recognized and controlled. Most critical is that, since the SLM does not produce full 2π modulation, there is always a dc component of the modulated light that is focused onto the optical axis. In order to separate the dc peak from the designed diffraction pattern, we add a phase ramp to the desired modulation. Since the SLM cannot produce all values between 0 and 2π, multiple harmonics in addition to the fundamental of the designed diffraction pattern are produced. Although use of a carrier frequency in practical systems leads to an undesirable reduction in the usable bandwidth of the SLM, for our purpose of demonstration, this condition is acceptable. Before presenting the results of the experimental demonstrations, we review the method of pseudorandom encoding and describe how the modulator is configured for the experiments.

2. Review of the Method

The encoding procedure is designed for SLM's composed of one- or two-dimensional arrays of pixels. However, the subsequent equations describing the procedure [in particular, Eqs. (4)–(6)] are presented only as a function of one spatial coordinate. This is done in part to simplify presentation, but more so to emphasize that the performance of the procedure depends directly on the total number of pixels. [The generalization to two dimensions simply requires the replacement of any spatial coordinate x or spatial frequency component f_x with the two-dimensional components (x, y) or (f_x, f_y).] The encoding procedure can be applied to any array of arbitrarily positioned pixels, but in the experiments we considered only regularly spaced arrays. The pixels have identically shaped apertures. Each pixel has unity-amplitude transmittance and can produce any value of phase between 0 and 2π. For purposes of design the value of the phase of the ith pixel located at position x_i is assumed to be its statistical average:

$$\bar{\psi}_i = \langle \psi_i \rangle, \tag{1}$$

where $\langle\ \rangle$ represents the expected-value or ensemble-average operator. The value of the phase actually produced by the pixel is

$$\psi_i = \bar{\psi}_i + \delta\psi_i, \tag{2}$$

where $\delta\psi_i$ is an unbiased random-phase offset. The offset is selected from a prespecified random distribution. The distribution determines the design value of the effective amplitude produced by the pixel. The set of all uniform random distributions with spreads between 0 and 2π is especially convenient for this purpose. Amplitude control between unity and 0 is provided according to

$$\bar{a}_i = \langle \exp(j\delta\psi_i) \rangle = \operatorname{sinc}\left(\frac{\nu_i}{2\pi}\right), \tag{3}$$

where ν_i is the spread of the uniform distribution. Since the values of Eq. (3) are in the range between

zero and one, we always assume that the full-complex modulation is normalized so that its maximum amplitude is unity. (A maximum amplitude of less than one would reduce diffraction efficiency and increase noise.)

To summarize the algorithm, we set the design or effective complex modulation $(\bar{a}_i, \bar{\psi}_i)$ equal to the desired full-complex modulation. Equation (3) is inverted to solve for spread ν_i. A value of $\delta\psi_i$ is selected from a uniform unbiased distribution of spread ν_i by a random number generation routine. The random-phase offset is added to the design value of phase [see Eq. (2)] to produce the actual modulation $(1, \psi_i)$.

The complex transmittance produced by a phase-only SLM consisting of an array of N identical subapertures centered at positions x_i is

$$t(x) = \sum_{i=1}^{N} \exp(j\psi_i) r(x - x_i), \tag{4}$$

where $r(x)$ is a function defining the clear area of the subaperture. The effective complex transmittance resulting from encoding each pixel is likewise written

$$\bar{t}(x) = \sum_{i=1}^{N} \bar{a}_i \exp(j\bar{\psi}_i) r(x - x_i). \tag{5}$$

The quality of the diffraction patterns produced by pseudorandom encoding can be appreciated by consideration of the expression for the expected intensity of the diffraction pattern. It has been shown to be[1]

$$\bar{I}(f_x) = |\bar{T}(f_x)|^2 + N(1 - \eta)R^2(f_x), \tag{6}$$

where f_x is the spatial frequency, $\bar{T}(f_x)$ and $R(f_x)$ are Fourier transforms of $\bar{t}(x)$ and $r(x)$, respectively, and the average intensity transmittance

$$\eta \equiv \frac{1}{N} \sum_{i=1}^{N} \bar{a}_i^2 \tag{7}$$

can be viewed as a type of diffraction efficiency, as we describe below.

Equation (6), the expected value of the intensity, is composed of two terms. The first term represents the specular component of the diffraction pattern (caused by the coherent superposition of wave fronts originating from the SLM). It is identical to the diffraction pattern of the desired full-complex modulation, i.e., the square magnitude of Eq. (5). The second term represents the diffuse component (or speckle pattern) that is due to the superposition of randomly phase wave fronts. Its spatial distribution is identical to the diffraction pattern from the aperture of a single pixel. Many diffraction patterns of interest are much more directional than the pattern of a diffuse scatter. This directionality gain over the more nearly white background noise often makes it possible to produce a close approximation to the desired diffraction pattern.

The amount of noise, and thus the quality of the reconstruction, is largely understood in terms of the diffraction efficiency η. The diffraction efficiency indicates the fraction of energy that appears in the coherent portion of the diffraction pattern. Since phase-only modulation does not attenuate, the remaining fraction of the incident energy, $1 - \eta$, is diffracted into the incoherent speckle background. In Ref. 1 we interpreted ηN as an effective number of nonrandom phase-only pixels. Equation (6) shows that increasing the diffraction efficiency (or equivalently, the number of nonrandom pixels) reduces diffuse scatter and would thus be expected to produce more accurate designs. Another way of viewing the relationship between the desired modulation and its pseudorandom encoding is that the more efficient a full-complex modulation is, the greater is its similarity to phase-only modulation.

An expression for the standard deviation of the intensity of the diffraction pattern is also presented in Ref. 1. It provides a precise statistical bound on the accuracy of the diffraction pattern at every point across f_x. Although the expression is not duplicated in this paper, it is used here to calculate error bars on plots of the designed intensity patterns. Although the effective number does not provide as detailed information as the standard deviation, it can be calculated with much lower computational overhead. Also it is well-enough correlated with the standard deviation that it can be used to gauge the quality of the diffraction pattern in real-time applications. Standard deviation, η, and N can be used together to define peak-to-background noise and the signal-to-noise ratio as was done in Ref. 1.

3. Experimental Configuration

The experiments are performed with two LCTV panels from the red channel of two different TVT6000 video projectors. Oscilloscope measurements show that the signals from the red, green, and blue driver boards of the projector are essentially identical when a black-and-white signal is applied to the video input. One LCTV is electrically reconnected to the driver board for the red channel, and the other LCTV is connected to the driver board for the blue channel (as illustrated in Fig. 2). Each of the two panels is connected through its own cable extender that we fabricated from a 1-m, 23-conductor, flat-laminated cable from Parlex Corporation and two zero-insertion-force flexible-printed-circuit connecters from Amp Incorporated.

The green LCTV is left in the projector. This is used to preview visually the image exactly as it is displayed by the LCTV panels. This provides an especially simple way to verify that the LCTV's respond correctly to the applied image. In fact, when a RS-170 standard composite video is applied to the video input of the projector, we observe that

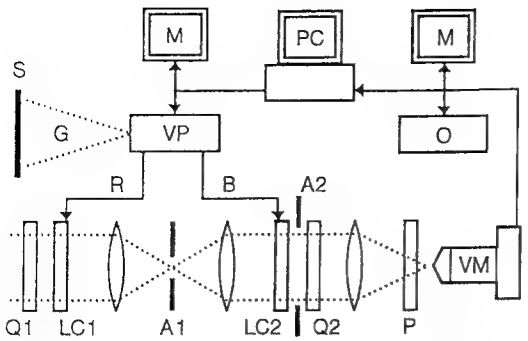


Fig. 2. Apparatus used for the experiments: VM; video microscope (or optical powermeter where noted in text); PC, personal computer that contains a frame grabber; VP, video projector; LC1, LC2, LCTV's from the video projector; S, viewing screen; M, video monitor; O, digital oscilloscope; P, polarizer; A1, A2, apertures; Q1, Q2, quarter-wave plates.

only 440 video lines and only ~94% of the viewable portion of each line is displayed by the projector. In terms of sending a 640 × 480 digitized image in video format from our frame grabber (Dipix 360f with optional video-display board), only 600 × 440 pixels are displayed. Since the LCTV has 480 × 440 pixels, there is a 1:1 mapping from frame-grabber lines to LCTV lines and a 5:4 mapping between pixels on a line.

A second issue in the mapping of the digital pixels is the modulation transfer function of the display. Loudin noted that a single video line produces two lines of video on the projected image from the LCTV.[10] We also observed this. We further note that the projected lines are roughly half the brightness possible. When two adjacent rows are turned on, one fully bright row is displayed, and there is a dim row to each side of the bright row. Apparently, the LCTV pixels cannot fully charge during the pixel time slot, and thus the sampling time (of the thin-film transistors) is extended to overlap with that of the adjacent lines. With projector sharpness set at minimum the effect is of roughly the same magnitude in both the horizontal and the vertical directions.

The optical apparatus used to produce and measure diffraction patterns is shown in Fig. 2. A collimated, nearly uniform-intensity laser beam of 632.8-nm wavelength and linear polarization illuminates the entrance port of the optical system. Standard optical components between the laser source and the entrance port are used to filter and expand the beam spatially, rotate the polarization, and control the intensity of the beam. The SLM consists of the components between the two quarter-wave plates. Light from the output of the second wave plate is focused with a 381-mm focal-length $f/5$ (Sorl F15/5) lens. The optical Fourier transform of the SLM is observed with a video microscope that consists of a Cohu 4915-2000 CCD video camera with an active imaging area of 6.4 mm × 4.8 mm, a C-mount 120-mm extension tube/microscope barrel, and a microscope objective. The camera output is observed on a video monitor or captured with the frame grabber. The frame grabber is configured to simultaneously acquire images while continuously outputting a single image. Individual video lines are viewed with a digital oscilloscope (HP54503A) that has built-in video triggering.

Alignment of the two LCTV's. The first LCTV (LC1) is imaged onto the second one (LC2) with a pair of 228.6-nm focal-length, $f/4.5$ imaging lenses (Plummer lenses purchased from **MWK**). The second LCTV is rotated by 180° to account for the image inversion produced by the $4f$ imaging system. An Ealing adjustable rectangular aperture (A1) at the Fourier plane between the two lenses is set to pass only the signal information surrounding the central diffraction order of the LCTV grating/pixel pattern. The coordinates used to describe the polarizations are chosen so that x, or horizontal, is in the direction of the video scan line, y is in the vertical direction of the video, and z is opposite of the direction of laser propagation. Both LCTV's are oriented so that the laser light emerges from the side of the panels on which the electrical cable is visible. The first LCTV (LC1) is oriented so the cable is attached below the horizontal axis. In this coordinate system the first quarter-wave plate (Q1) is illuminated from a red helium–neon laser with a linear polarization of 70° from x and toward y. The fast axis of the first wave plate is at 45°, and the fast axis of the second quarter-wave plate (Q2) is at −45°. The polarizer (P) is oriented to pass light at 70°, the same as the incident linear polarization. The polarizer produces a substantial amount of wave-front distortion across its full aperture. The illuminated area and consequently distortion are reduced by placement of the polarizer near the focus of the Fourier-transform lens. The projector brightness and contrast controls are set to maximum, and the sharpness is set to minimum.

The effect of maximum sharpness is clearly evident on broadly extended diffraction patterns. In particular, if the LCTV is driven by a digital image of white noise, the envelope of the resulting speckle pattern resembles the diffraction pattern of a single pixel except that it is modulated in the horizontal direction by six high-contrast stripes covering the separation between adjacent diffraction orders of the SLM. This scallop-shaped pattern is reduced for the minimum sharpness setting. We further reduced the effect in our experiments by programming multiple adjacent pixels with the same video level, i.e., grouping pixels together into superpixels. The scalloped diffraction pattern is probably due to filtering produced from the combination of sharpness control, the LCTV pixel integration/charging time, and the 5:4 mapping between the frame grabber and the LCTV pixels.

Aperture A2 is chosen to be 15.2 mm × 15.2 mm. This limits the illuminated area of the SLM to 272 pixels by 332 lines. For the available lenses this

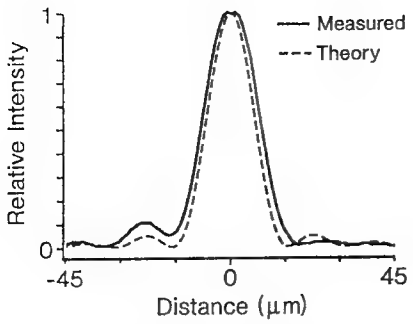


Fig. 3. Point-spread function of the SLM aperture as measured and predicted. The cross section is taken across the center of the spot perpendicular to the horizontal lines of the SLM.

permits registration of the image of the first LCTV onto the second LCTV to better than 0.5 pixel in both x and y. We verified the alignment by modifying the optics to make amplitude modulation evident. We did this by rotating the wave plates into an appropriate orientation and placing the polarizer and the video microscope immediately after the second LCTV. Aperture A1 at the Fourier plane between the two LCTV's is also opened so that the edges of individual pixel apertures from the first LCTV are observable through the pixel apertures on the second LCTV. Single line and row patterns are programmed onto the LCTV's, and micropositioners are used to align the patterns. The microscope is mounted on translation stages, and the alignment is checked at several positions across aperture A2. The alignment errors only become noticeable close to the edges of the aperture and appear to be caused by a slight barrel distortion in the imaging optics. After the alignment is set and/or tested in this way, aperture A1 is once again set to pass only the central diffraction order, and the system is reconfigured according to Fig. 2.

In addition to registration, aperture A2 is also needed in order to improve optical flatness so as to achieve near-diffraction-limited Fraunhofer patterns. We determined this by measuring the point-spread function of the SLM aperture with the video microscope. For the modulated light to be separated from the unmodulated light, the SLM, rather than being programmed with a constant gray-scale level, is programmed with a periodic ramp that displaces the point-spread function from the optical axis. (Details on selection of the ramp function and offset are given in Section 4.) Figure 3 compares the spread function measured at the focal plane of the transform lens with the diffraction pattern ideally produced by a 15.2-mm aperture and a 381-mm transform lens. The experimental intensity trace is taken across the center of the intensity peak, and it is aligned along a horizontal line of the camera in order to improve resolution. The correct scale for the spatial coordinates was determined by the manufacturer-quoted dimensions for the CCD imager together with the measured magnification of 9.76X (nominally a 10X objective is used) for the video microscope.

Amplitude-modulating properties of the SLM. We also measured (using a photodetector and a powermeter in place of the video microscope) the variation in intensity of this central diffraction order when the uniform gray-scale image from the frame grabber is varied from full black to full white. The residual amplitude modulation is ±3.7% of the average amplitude modulation. The maximum transmittance is at a gray scale of ~128, and there are two local minima at ~40 and ~256.

We also measured energy utilization of the SLM between the input face of Q1 and the output face of Q2. We did this by measuring the intensity flux across the (effective) input and output apertures (A2 projected onto Q1 and Q2.) With the SLM set for maximum transmittance, only 0.325% of this incident energy is measured at Q2. Since one desirable feature of the eigenpolarization mode is its low loss, we were initially concerned about the low optical efficiency. We also configured the system as a phase-only modulator by illuminating each LCTV with a linear polarization and passing the linear polarization that most closely approached phase-only operation. The efficiency of this cascade is 1/6 that of the eigenpolarization arrangement. (It also has a similar phase-modulation range.)

The most significant loss factor is apparently related to the shadow mask in each LCTV panel. Measurements under a microscope indicate that the pixel aperture is roughly 28 μm × 28 μm on a pitch of 56 μm horizontally by 46 μm vertically leading to a fill factor of approximately 30%. Therefore only 30% of the energy is transmitted through the first shadow mask. Based on the Fourier-series analysis of a square wave, only 30% of this energy remains in the zero order at the filter plane between the two lenses of the imaging system. Owing to the spatial filtering of the nonzero orders, the second LCTV should produce an identical reduction in the transmitted flux. Overall, these losses would lead to a transmittance of 0.86%. Thus in this configuration even moderate fill factors can be the dominant source of loss.

Phase-modulating properties of the SLM. We initially measured the phase shift by building a Mach–Zehnder interferometer around the SLM arrangement shown in Fig. 2. The reference beam was tilted to produce fringes vertical to the video scan. The shift of the fringe between the row driven with gray scale 0 and a row driven with another value of gray scale was measured. These measurements indicate that the phase increases monotonically from 0° to 326°. A rough curve that approximately describes the phase shift as a function of gray scale is a curve that has a linear slope of 2° gray-scale level between gray-scale levels 0 and 140 and 0.4° per gray-scale level between 140 and 255.

This measurement procedure, however, appears to be subject to several errors, including coherent noise,

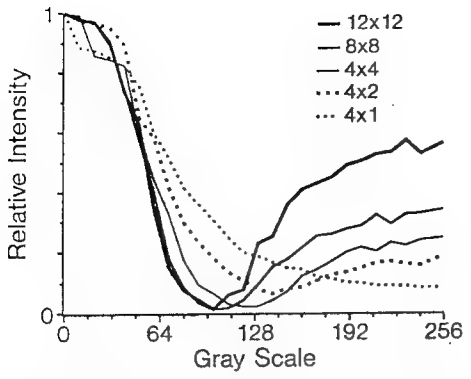


Fig. 4. On-axis diffraction intensity for random binary gray-scale patterns and with superpixels of various sizes. Each curve is normalized to the peak intensity with the gray scale equal to zero.

curvature, low contrast, and vibration (caused by air turbulence) of the fringes. Furthermore, we had concerns about errors introduced by the additional optical components of the interferometer. A second measurement procedure was devised in order to confirm the phase measurements. The procedure also indicates the dependence of the phase modulation on the spatial frequencies of the signal applied to the SLM. To perform the measurement, one loads a pseudorandom binary-level pattern into the frame grabber. An aperture and a powermeter are positioned at the focal plane of the transform lens shown in Fig. 2 so as to intercept the central diffraction order. Half the SLM pixels are randomly selected and driven with a gray-scale value of 0. The other half are driven with a second gray-scale value chosen between 0 and 255. The intensity measured on the powermeter is recorded as a function of gray scale. The measurement was repeated for various SLM pixel sizes, where an SLM pixel is defined to be an $n \times m$ array of LCTV pixels that are programmed with the same gray-scale value. These results are plotted in Fig. 4.

Diffraction from a binary-level, phase-only modulation can be understood by the analyses from Ref. 1 that were used originally to describe pseudorandom encoding. In this case each pixel is programmed to have an identical effective amplitude that produces an expected intensity pattern consisting of a diffraction-limitd spot centered on the optical axis and a broad background noise level. Since the light from all SLM pixels mutually interferes at the intersection of the optical axis and the focal plane (i.e., f_x equal to zero), as illustrated in Fig. 1, the intensity of the spot depends directly on ψ, the difference between the two phases present on the SLM. The phasing on axis can be evaluated by deterministic analysis alone. For an ideal phase-only SLM in which each pixel can be controlled independently of all others, the intensity of the diffraction peak on the optical axis is

$$I(0) = N^2R^2(0)\frac{1 + \cos\psi}{2}, \tag{8}$$

where N is the number of SLM pixels and $R(f_x)$ (introduced in Section 2) is the element factor corresponding to the pixel clear aperture. As would be expected, Eq. (8) indicates that there is perfect cancellation of the spot for a phase difference of 180°. (In practical measurements a finite detector size and nearby speckles would introduce errors into the measurement of power. However, since N is usually a large number, speckle noise is not a serious source of error over most of the phase range. Assuming that the SLM has adequate phase range, it would also be possible to offset the lower gray-scale value from zero to a larger value in order to more clearly measure the 180° point.)

Based on the initial phase measurements with the Mach–Zehnder interferometer, it is reasonable to expect that there would be a simple and direct relationship between the phase difference and the nonzero gray-scale value. Indeed, the curves in Fig. 4 show a qualitative agreement with this model in that they resemble a cosine function for gray-scale values of <128. For gray-scale values of >128 the intensity increases but much more slowly than for the lower gray-scale values. This tracks with the interferometer measurements of phase in which the phase sensitivity to gray scale decreased at large values of gray scale.

The curves indicate that the phase-modulation range increases with SLM pixel size. The maximum range in these curves is for the 12 × 12 array. The maximum value of $I(0)$ after the phase exceeds 180° is 0.57. Inverting Eq. (8) gives an estimated modulation range of 278°. For the 4 × 1 SLM pixel there is barely a 180° range. The correspondence with the original interferometer measurements appears to be good for low spatial frequency patterns of modulation; however, the actual phase modulations needed for the encoding experiments have a range of frequencies from high to low.

Summary of SLM properties. The cascade of two SLM's was originally intended to produce a phase modulator with a range of 360°. The measured SLM does have a small degree of residual amplitude modulation. It also has a large phase-modulation range that is close to 360° but only for modulations with low spatial frequencies. The phase-modulating characteristics can be made independent of spatial frequency by use of large SLM pixels; however, this significantly reduces the number of available pixels for the encoding experiments. (For the case of 12 × 12 SLM pixels and the available clear aperture there are fewer than a thousand invidiaul pixels.)

4. Experimental Demonstrations of Pseudorandom Encoding

Adjustments made for demonstrations with the available SLM. Since the available SLM does not achieve

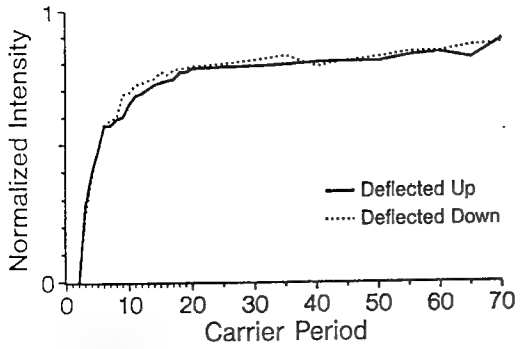


Fig. 5. Intensity of the deflected point-spread function as a function of the carrier period. Intensities are normalized to the intensity of the undeflected (on-axis) point-spread function. The unit for the carrier period is the number of horizontal SLM lines per period.

a 360° range, there is usually a dc component in the spatial modulation and a corresponding diffraction peak at f_x equal to zero. We chose to separate the dc component by multiplying the modulation by a periodic function. We accomplish this by adding together a linear ramp function with the gray-scale values of the encoded signal, **modding** the result to between 0 and 255, and placing this image in the display memory of the frame grabber. Several harmonically related diffraction patterns are generated, of which the fundamental is usually the brightest. It is this fundamental diffraction pattern that we compare with those anticipated from our presumed phase-only modulator.

The fundamental spatial frequency of the phase ramp (i.e., the carrier frequency) also affects the quality of the diffraction pattern. This is most evident in that the point-spread function decreases with increasing phase ramp or carrier frequency. This is illustrated in Fig. 5, in which diffraction efficiency is plotted against the period of the carrier. Efficiency for this measurement is defined as the peak intensity normalized by the intensity of the spread function when it is centered on the optical axis (specifically, when all pixels of the SLM are programmed to gray-scale value 0). The peak intensities are measured (by the digital oscilloscope in Fig. 2) across a single video line of the CCD camera (rotated 90° from horizontal lines of the SLM). Figure 5 can be viewed as a type of modulation transfer function. The curve shows for periods between 20 and 70 SLM lines that the diffraction efficiency is close to flat. The efficiency falls off rapidly at <10 lines. For the pseudorandom-encoding demonstrations, presented below, a period of 20 lines was found to be acceptable. As the period increases to >20 lines, interference from overlap with the dc and the first-harmonic order becomes increasingly significant. Figures 5 and 4 indicate that individual pixels are not dependent of the modulation of nearby pixels. It is surprising that there is still residual correlation between pixels that are 10 and 20 lines apart. Certainly, by any criteria, the resolution of the SLM is far worse than one pixel.

The various issues relating to making the SLM behave as an array of independently controllable pixels cannot be well satisfied with the given constraints of maintaining a diffraction-limited aperture, the low spatial resolution of the SLM, the lack of 1:1 mapping between display board pixels and SLM pixels along the scal line, and the need for a large number of pixels. These constraints led us to program the SLM as clusters of superpixels. Each superpixel consists of 4 pixels along a single scan line for 22,000 total pixels, which is roughly the same number of pixels (16,384) as used in the simulations in Ref. 1. Using a 4 × 2 superpixel was not considered after noting, based on Fig. 4, that the total phase-modulation range would increase to only 193°, an increase of as little as 13°. Thus the SLM is programmed as if it consists of 68 × 332 superpixels, and the carrier has a period of 20 lines in the vertical direction.

Design and encoding of full-complex modulations for beam shaping. Three apodizations are designed to produce rectangularly shaped diffraction patterns centered on the optical axis. The apodization is the product of a sinc function and a Dolph window in the x, or horizontal, direction.[11] This window is chosen to reduce the Gibbs ripple that can result from the finite extent of the SLM truncating the sinc function. The specific Dolph function selected is defined by the sidelobe level of its Fourier transform $d(f_x)$, which for these designs is 26 dB. Higher sidelobe-level Dolphs are known to further reduce the Gibbs ripple, but they also increase transition bandwidth. As for pseudorandom encoding, higher-sidelobe-level Dolphs are more steeply tapered, which reduces the value of diffraction efficiency η [see Eq. (7)] and consequently increases the speckle noise [see Eq. (6)]. In the computer simulations of random-encoding-designed diffraction patterns in Ref. 1, a 26-dB sidelobe level was empirically found to provide a good trade-off between reduction in the Gibbs ripple and background noise. The same apodization was applied both in the horizontal and the vertical directions. In the designs for the experiments presented here, the apodizations vary in only the vertical direction. The one-dimensional apodization is used with the objective of further increasing the diffraction efficiency and reducing the random noise inherent to the encoding procedure. The apodizations are chosen to produce diffraction-intensity patterns of the form

$$I(f_x, f_y) \propto \mathrm{sinc}^2(f_x)[d(f_y) \overset{*}{\wedge} \mathrm{rect}(f_y/w)]^2, \qquad (9)$$

where $\mathrm{sinc}^2(f_x)$ is the point-spread function in the x direction and $d(f_x)$ is the Fourier transform of the Dolph window. The amplitude in the y direction is designed to approximate a rect function of width w for values of 2, 3, and 4. The width is normalized in terms of the unity width (from peak to first null) of

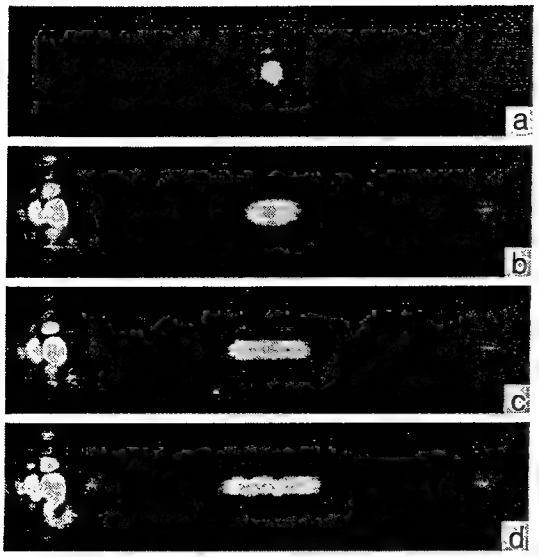


Fig. 6. Gray-scale images of the measured diffraction-pattern intensity: (a) the measured point-spread function (cross section shown in Fig. 3). The measured diffraction patterns that are designed to approximate a rect function in f_y are of widths (b) w = 2, (c) w = 3, and (d) w = 4. The images are oriented so that f_y is horizontal to the page.

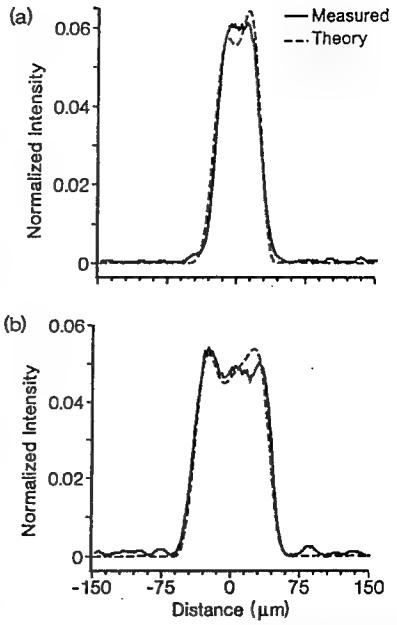


Fig. 7. Theoretical and measured diffraction patterns. Measured intensity curves are normalized with respect to the peak intensity of the measured point-spread function in Fig. 3. Cross sections are taken from the centers of the diffraction patterns shown in (a) Fig. 6(b) and (b) Fig. 6(c). Theoretical intensity plots are scaled by 1.96× in (a) and by 1.53× in (b) with respect to the peak intensity of the theoretical point-spread function in Fig. 3. Both measured curves are plotted with a +4-μm offset from the center of the measured point-spread function.

$\text{sinc}^2(f_y)$, the point-spread function in the y direction. From Fig. 3 it can be seen that $w = 1$ corresponds to a physical distance of ~15 μm.

Each of the three apodizations are pseudorandom encoded by Eqs. (1)–(3). As described above, these modulations are then multiplied by a carrier frequency with a period of 20 video lines before being loaded onto the frame grabber. Theoretical diffraction patterns of the apodizations are also calculated. This is done by a 256 × 4096 fast Fourier transform on the 68 × 332 designed modulations (zero padded to the size of the fast-Fourier-transform array). Cross sections in the y direction from the centers of the theoretical diffraction patterns are saved for comparison with the measured diffraction patterns.

Summary of experiments and their comparisons with theory. The diffraction patterns are recorded with the CCD camera in the video microscope rotated by 90° from the x (horizontal) and the y coordinates of the SLM. The resulting images are shown in Fig. 6. In each of the four images the CCD camera is saturated by roughly the same factor with respect to the peak intensity of the desired pattern in order to aid in comparisons of the background levels. Figure 6(a) is the image of the point-spread function. (The cross section of this diffraction pattern is shown in Fig. 3.) The cross sections corresponding to Figs. 6(b)–6(d) are shown in Figs. 7 and 8(a). Except for changes in attenuator settings, the corresponding curves and images were recorded under identical conditions and within a few minutes of each other.

Figure 6(a) shows that a substantial amount of energy is deflected into the fundamental order and that there is much less energy at dc and the first harmonic. The first diffraction order (not shown) to the left of dc is much weaker than the fundamental but stronger than dc and the first harmonic. Figures 6(b)–6(d) show the successive widening of the beam footprint. The first harmonic is still quite faint for the shaped beams, but dc is now very bright. As can be seen from the vertical axes in Figs. 7 and 8(a), the intensity of the fundamental is much less than it is in Fig. 6(a). The sidelobe structure of the dc component in Figs. 6(b)–6(d), although more heavily saturated, closely resembles the point-spread function in Fig. 6(a). Each of these diffraction patterns also shows, as expected, both a background speckle pattern and sidelobes along the narrow direction of the beam footprint.

Figures 7 and 8(a) provide more precise information describing the measured diffraction patterns. The three measured curves are normalized to the peak intensity of the experimental point-spread function in Fig. (3). The three theoretical curves were initially normalized by the peak intensity of the theoretical curve in Fig. 3, but the curves were found to be somewhat lower than the experimental intensities. Multiplying the theoretical curves by vertical scale factors (between 1.53 and 1.96 as noted in the corresponding figure captions) helps to show just how similar the curves are in shape. Offsetting the measured data by 4–5 μm further improves the correspondence, but to a much lesser degree. In

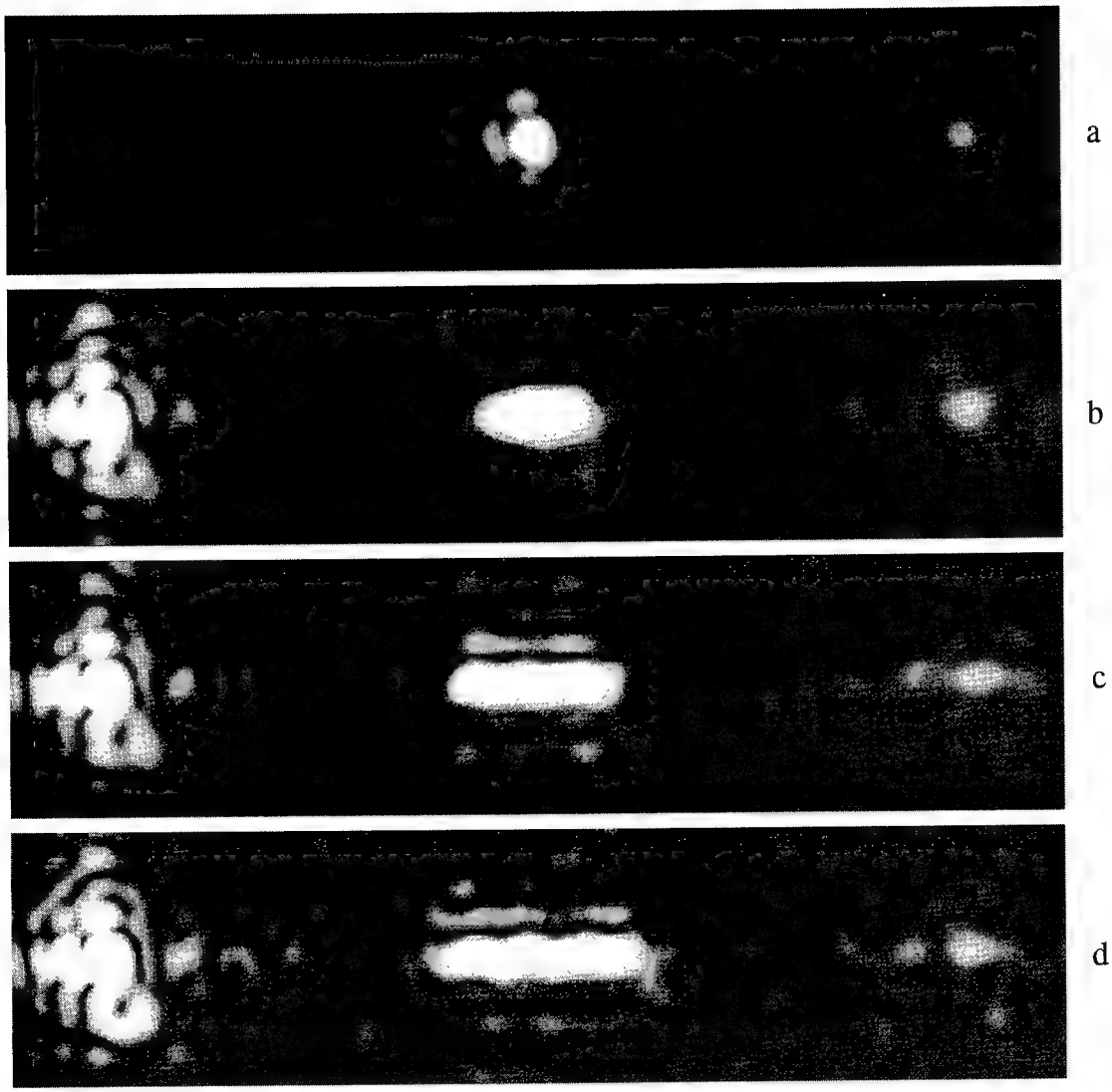


Fig. 6. Grayscale images of the measured diffraction pattern intensity. a) The measured point spread function which would ideally be of the form sinc(x)sinc(y). The measured diffraction pattern designed (using a window design procedure) to approximate b) rect(x/2)sinc(y), c) rect(x/3)sinc(y), d) rect(x/4)sinc(y). Each image is saturated by approximately the same factor with respect to the peak intensity of the (desired) pattern centered at the carrier frequency.

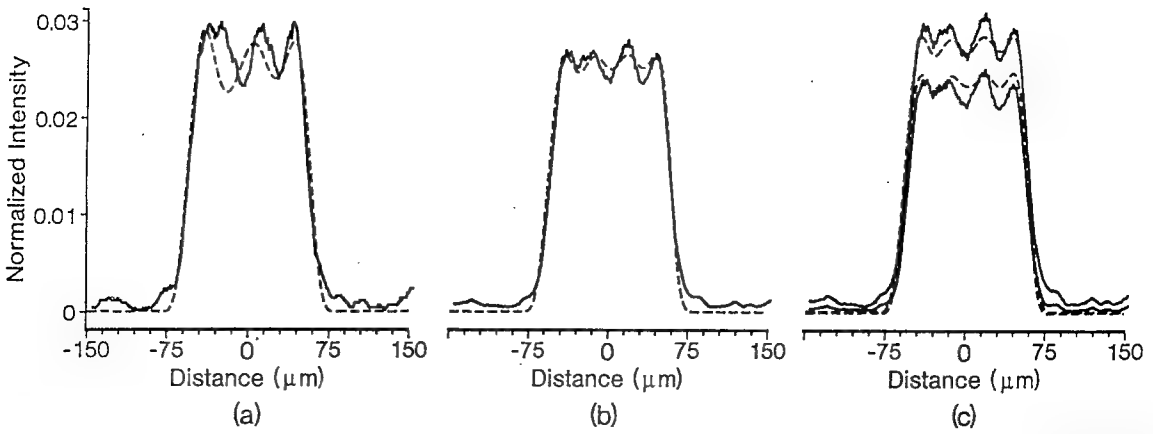


Fig. 8. Theoretical (dashed curves) and measured (solid curves) diffraction patterns: (a) cross section of the center of the diffraction pattern shown in Fig. 6(d); (b) the average of 120 cross sections with a different random seed for each experiment; (c) the error bars (plus- and minus-one standard deviation) of the average intensity in (b). Curves are normalized the same as in Fig. 7 except that the measured curves are plotted with a +5-μm offset and the theoretical curves are scaled in intensity by a factor of 1.64×.

these experiments the scale-factor difference between experiment and theory is the most significant discrepancy. The offset in position is rather small when compared with the diameter of the point-spread function. These discrepancies are probably due to one or more of the nonideal properties of the actual SLM, none of which were factored into our theoretical model. It is not our goal to account for these differences quantitatively but rather to gain insight into how the encoding method would behave when 360° modulators become available. For this reason we think that the degree of correspondence, while qualitative, is nonetheless excellent, and we are encouraged by this result.

Further evidence of the correspondence between experiment and theory can be demonstrated by repeated trials of the pseudorandom-encoding procedure. An example of this is given in Fig. 8. Here, the experiment leading to the result in Fig. 8(a) is repeated 120 times. The only difference between each trial is that the random-number generator is initiated with a different random seed for each encoding. The average intensity of the 120 trials (both experiment and theory) is calculated, as is the plus- and minus-one standard-deviation error bars of the intensity pattern. The number of trials is adequate to produce theoretical diffraction patterns that, for purposes of plotting, are nearly identical to the ensemble average that would be found for an infinite number of trials (i.e., found by Eq. (6) and the equation for standard deviation in Ref. 1). One can appreciate the small difference by noting that the theoretical curve in Fig. 8(b) has almost equal-level ripples, whereas the expected intensity for an infinite number of trials has exactly equal-level ripples. This degree of agreement between theory for a finite and theory for an infinite number of trials was found for the theoretical results in Fig. 8(c) as well.

The measured and the theoretical average curves in Fig. 8(b) show more clearly the correspondence than does Fig. 8(a). The discrepancies between the two average curves indicate differences caused by systematic errors in the experiment.[12] Most notably, the measured curve exhibits more ripple and higher sidelobes, or noise. The larger error bounds in Fig. 8(c) indicate that more random noise is being introduced into the measurement than that for the theory (in which noise is due only to the random statistics used in encoding). The measured error bounds are only ~60% wider than theoretical in-band. Nonetheless, we think that the degree of agreement is quite good considering that the SLM is so different from the assumed device, a 360° phase-only SLM. Most important is that the diffraction patterns produced do qualitatively agree with our theory in that the beam shape is controllable and there is a broad, near-uniform intensity speckle background. As devices that do produce 360° become available, it appears that it will be possible to synthesize diffraction patterns with greatly improved accuracies. Elimination of the need for a carrier frequency will also make it possible to use the entire bandwidth of the SLM for signal processing.

5. Summary and Conclusions

SLM measurements and characterization procedures. A secondary goal of this paper has been to describe the properties and the characterization of the SLM. In most previous studies of electrically addressable SLM's, the modulation transfer function has been ignored. Our measurements, in which we applied different carrier frequencies, indicate that spatial frequencies of <5 SLM lines are substantially filtered. There is even a nontrivial amount of filtering at 10–15 SLM lines. The filtering reduces the phase-modulation range and causes the random-phase components at nearby pixels to become statistically correlated. This point may have been overlooked in some previous interferometric studies of SLM phase-

modulating properties in which large blocks of pixels were programmed with the same gray-scale value.

Recently, two groups have focused on common-path interferometric approaches in order to eliminate vibrations.[13,14] One procedure is as follows: (1) the beam is split with a Ronchi ruling; (2) one beam is passed through a reference portion of the SLM (programmed to gray scale 0), and the second beam is passed through a portion of the SLM that is modulated with a second gray-scale value; (3) the beams are recombined and detected by camera; and (4) the phase shift of the sinusoidal interference pattern is measured.[13] The second procedure is as follows: (1) a 50% duty-cycle square wave (Ronchi ruling) is programmed onto the SLM with half the gray scale values set at 0 and the other half set to a second level; (2) the diffraction pattern of the SLM is detected by camera; and (3) the phase is decoded based on the relative intensities of the first and the third diffraction orders.[14] An alternative common-path procedure is described here in which a pseudo-random binary-phase-modulation pattern is applied to the device. The method is useful in that it is an *in situ* measurement procedure that requires no extra beam splitters or combiners, and various-sized superpixels can be evaluated. Since only the intensity of the on-axis diffraction pattern is needed, the measurement can be performed with a single photodetector. Also, over most of the phase range, the diffraction-limited spot is much brighter than the background speckle noise, which can lead to repeatable measurements of phase. For SLM's that have fairly uniform characteristics across the surface of the device and for which the amplitude variation is also known, it will also be possible to measure the phase with high precision.

Experimental demonstration of pseudorandom encoding. We attempted to develop a 360° modulator by cascading two TVT6000 SLM's. A modulation range of 326° was measured initially. However, the phase modulation was also found to depend on the spatial frequencies present on the SLM. For high spatial frequencies the modulation range decreases to ~180°. Although it was found that increasing the size of superpixels also increases the phase-modulating range, the superpixels could not be made large enough to simultaneously eliminate the sensitivity of the SLM to spatial frequencies and provide a reasonably large number of pixels for the encoding experiments. It is comforting to note that the carrier frequency and the apodization are slowly varying functions. However, the random-phasor portion [see Eq. (2)] of the modulation pattern (that controls amplitude) contains the highest spatial frequencies. The portions of the apodization that are designed to produce the smallest amplitudes will have the largest random spread [see Eq. (3)] and thus the most rapid phase transitions between adjacent pixels. Thus the low-amplitude portions are expected to be the most in error owing to the modulation-transfer-function limitations of the SLM. In addition, the nonlinear mapping between gray scale and phase was not factored into the theoretical diffraction pattern. The use of the carrier frequency also means that different portions of the apodization pattern are phase encoded by different ranges of the nonlinear mapping curve. Despite these properties, which are significantly different from an ideal array of independent 360° phase-only pixels, qualitatively similar results were produced. Not only were the diffraction patterns **very** similar in shape to ideal, but a broad background of speckle was observed. The statistically averaged experimental results were also quite similar to theory, and the error bounds were only somewhat noisier than theory. These experimental results lead us to conclude that pseudo-random encoding is a robust procedure in that the theoretical predictions will reasonably match the experimental results without use of a perfectly ideal SLM.

This research is sponsored by Advanced Research Projects Agency through Rome Laboratory contract F19628-92-K0021, U.S. Army Research Office contract DAAH04-93-G-0467, and National Aeronautics and Space Administration cooperative agreement NCCW-60 through Western Kentucky University.

References

1. R. W. Cohn and M. Liang, "Approximating fully complex spatial modulation with pseudo-random phase-only modulation," Appl. Opt. **33**, 4406–4415 (1994).
2. J. L. Horner and P. D. Gianino, "Phase-only matched filtering," Appl. Opt. **23**, 812–816 (1984).
3. N. Konforti, E. Marom, and S.-T. Wu, "Phase-only modulation with twisted nematic liquid-crystal spatial light modulators," Opt. Lett. **13**, 251–253 (1988).
4. J. Amako and T. Sonehara, "Kinoform using an electrically controlled birefringent liquid-crystal spatial light modulator," Appl. Opt. **30**, 4622–4628 (1991).
5. K. Lu and B. E. A. Saleh, "Theory and design of the liquid crystal TV as an optical spatial phase modulator," Opt. Eng. **29**, 241–246 (1990).
6. J. L. Pezzaniti and R. A. Chipman, "Phase-only modulation of twisted nematic liquid-crystal TV by use of the eigenpolarization states," Opt. Eng. **18**, 1567–1569 (1993).
7. C. Soutar and S. E. Monroe, Jr., "Selection of operating curves of twisted-nematic liquid crystal televisions," in *Advances in Optical Information Processing VI*, D. R. Pape, ed., Proc. SPIE **2240**, 280–291 (1994).
8. C. Soutar and K. Lu, "Determination of the physical properties of arbitrary twisted-nematic liquid crystal cell," Opt. Eng. **33**, 2704–2712 (1994).
9. R. M. Boysel, J. M. Florence, and W. R. Wu, "Deformable mirror light modulators for image processing," in *Optical Information Processing Systems and Architectures*, B. Javidi, ed., Proc. SPIE **1151**, 183–194 (1989).
10. J. A. Loudin, "LCTV custom drive circuit," in *Optical Pattern Recognition V*, D. P. Casasent and T. H. Chao, eds., Proc. SPIE **2237**, 80–84 (1994).
11. F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," Proc. IEEE **66**, 51–83 (1978).
12. R. W. Cohn and J. L. Horner, "Effects of systematic phase errors on phase-only correlation," Appl. Opt. **33**, 5432–5439 (1994).

13. C. Soutar, S. E. Monroe, Jr., and J. Knopp, "Measurement of the complex transmittance of the Epson liquid crystal television," Opt. Eng. **33,** 1061–1068 (1994).

14. Z. Zhang, G. Lu, and F. T. S. Yu, "Simple method for measuring phase modulation in liquid crystal televisions," Opt. Eng. **33,** 3018–3022 (1994).

Using pseudorandom phase-only encoding to approximate fully complex distortion-invariant filters.

Laurence G. Hassebrook, Michael E. Lhamon, Raymond C. Daley

University of Kentucky
Department of Electrical Engineering
453 AH
Lexington, KY 40506-0046

Robert W. Cohn and Minhua Liang

University of Louisville
Department of Electrical Engineering,
Louisville, KY 40292
Rwcohn01@Ulkyvm.Louisville.Edu

ABSTRACT

Any desired diffraction pattern can be produced in the Fourier plane by specification of a corresponding input plane transparency. Complex-valued transmittance is generally required but, in practice, phase-only transmittance is used. Many design procedures use numerically intensive, constrained optimization. We, instead, use a non-iterative procedure that directly translates the desired, but unavailable, complex transparency into an appropriate phase transparency. At each pixel the value of phase is pseudorandomly selected from a random distribution whose standard deviation is specified by the desired amplitude. We apply the pseudorandom phase-only encoding to Hybrid Composite filter design. These filters are used in a filter bank architecture to perform intensity- and distortion-invariant pattern recognition.

2. INTRODUCTION

This study is motivated by a desire to design phase-only composite filters with a small amount of electronic computation and thereby permit programming of arbitrary spatial modulation at real-time rates. Popular design procedures (e.g. the Dammann grating,[1-3] simulated annealing,[4] iterative constrained optimization,[5,6] and other iterative procedures) are only practical if performed off-line due both to the numerical cost of performing Fourier transforms repeatedly and the further cost of evaluating the sensitivity of the transform with respect to a large number of pixels (frequently every pixel of the input plane spatial light modulator.) While it can be argued that for some applications all necessary phase-only filters can be computed off-line and stored in memory, in other applications either too much memory is needed to do this affordable or there may not be advance knowledge of what filter is needed.

There are many procedures in the area of computer generated holography, especially kinoforms, that permit direct synthesis of the input plane. These presuppose that the Fourier transform pair between the fully complex-valued input and Fourier planes are known and work by encoding the desired complex values to appropriate phase settings. The direct synthesis design procedures thus allow programming at real-time rates, if the desired Fourier plane pattern is known.[7] The amount of memory is also minimized if the complex valued Fourier transform pair can be written as an easily computed function.

Most frequently the direct procedures group a few adjacent pixels together as a single pixel that approximates several discrete settings over the complex plane (i.e., cell-oriented encoding).[8] This however reduces the space bandwidth, which is already quite small (typically 128 x 128 pixels) for current spatial light modulators, as compared to traditional fixed pattern holographic and diffractive optical elements. The procedure developed by Cohn and Liang,[9] which we apply here is also a direct method, but one where a continuous value of phase is selected for each individual

pixel independent of all other settings (i.e., point-oriented encoding).[8]

Cohn and Liang's method directly follows from mathematical models of far-field diffraction from arrays of randomly-phased point sources. More specifically, the phases are treated as independent and non-identically distributed random variables. Many previous studies have considered random phase diffusers in which the phase statistics are identically distributed across the diffuser. The objective of that work was improved holographic reconstruction of the intensity of objects and for which the phase was of no concern to the observer. In our design we are interested in reconstructing full complex objects from a variably-random phase-only filter plane. In our analysis and design procedure it is convenient to represent the random variables with a single probability density function that is varied in its mean and standard deviation. Our previous analysis[9], shows that on average, the pixels with phases drawn from distributions of larger standard deviation behave as if they have smaller amplitude transmittances. This result originally occurred to us when we noted an average contrast loss for arrays of phase-only pixels that have identically distributed phase errors.[10] We now recognize that this earlier result is just a special case of the more general result for non-identically distributed phases.

The ideal SLM, under consideration, modulates phase continuously over a full 2π range. Regularly spaced pixels are assumed in the examples but are not required in theory. These characteristics may be achievable with twisted nematic[11] liquid crystals. For binary and multi-level optics the quantized values of phase can be modeled as statistical departures from the desired analog phase and this added effect on the diffraction pattern can be estimated.[10,12]

3. RANDOM PHASE CONCEPT

Our design procedure specifies the degree of randomness (i.e. standard deviation) of phase at each pixel in order to approximate arbitrary values of amplitude. This is analogous to placing a diffuser of spatially varying roughness over the input plane. Increasing roughness at a pixel decreases its coherent contribution to the far-field pattern. The remaining incoherent, or diffused light is spread over the entire diffraction pattern and contributes a noise background that is frequently referred to as speckle. The design phase at any pixel is specified as the expected value of its random phase distribution. Thus the full complex input plane can be viewed as a cascade of a deterministic phase screen with a variable roughness phase screen.

In our design procedure, rather than average many trials, we instead select a single value of phase from the ensemble for each pixel. The expected complex amplitude and intensity of the far-field pattern for this situation is mathematically equivalent to that described in the previous paragraph. More important, the actual far-field pattern approximately resembles the average pattern that results from the coherent summation of a large number of random wavefronts. Thus, the formation of the far-field pattern can be viewed as a physical example of the central limit theorem.

3.1 Mathematical analysis of random phase

Definitions for arrays of pixels are used in the development of the statistical expressions. The complex transmittance of an individual pixel centered at position x equal to x_i in the input plane will be written as $a_i(x)$ and the transmittance of the array of the N individual pixels is

$$t(x) = \sum_{i=1}^{N} a_i(x) = \sum_{i} r(x-x_i)\,\exp(j\psi_i), \qquad (1)$$

where ψ_i is the phase shift produced by the i'th pixel of the SLM. The abbreviation r(x) = rect(x/w) has been used where w is the width of each pixel. The amplitude and phase of each pixel as shown in Eq. (1) has been defined in local coordinates centered around x=0 and then shifted to pixel locations x_i. In this and subsequent equations the inactive area between the pixels is treated as being zero amplitude. This is done in order to focus on the fundamental effect. The intensity of the far-field diffraction pattern is written as

$$I(f_x) = T(f_x)\,T^*(f_x) = \mathcal{F}\{\, t(x) \oplus t(x) \,\}, \qquad (2)$$

where $T(f_x)$ is the Fourier transform of the transmittance t(x) (i.e. $T(f_x) = \mathcal{F}[t(x)]$) and where $\oplus$ indicates the correlation

integral such that

$$a(x) \oplus b(x) = \int a(x'+x)\, b^*(x')\, dx'. \tag{3}$$

The squared intensity spectrum expressed in terms of the fourth-order autocorrelation of SLM transmittance is Unlike convolution, the order in which correlations are performed affects the result, and thus the brackets are required

$$I^2 = TT^*T^*T = \mathcal{F}\{[t(x) \oplus t(x)] \oplus [t(x) \oplus t(x)]\}. \tag{4}$$

in Eq. (4).

3.2 Statistical expressions

The expected value of a complex phasor of a random argument is frequently referred to as the characteristic function of the random variable ψ

$$M'(\omega) = \langle \exp(j\omega\psi) \rangle = \int_{-\infty}^{\infty} \exp(j\omega\psi) p_\psi(\psi)\, d\psi = 2\pi\, \mathcal{F}^{-1}\{p_\psi(\psi)\}, \tag{5}$$

where $\langle\ \rangle$ is our symbol for the ensemble average (i.e. expectation) operator and where $p_\psi(\psi)$ is the probability density function of ψ. The last equality in Eq. (5) emphasizes the well known Fourier transform relationship between the probability of the density function and the characteristic function. If the phases ψ_i are independent random variables then the expected value of pixel transmittance $a_i(x)$ is

$$\begin{aligned} \bar{a}_i(x) &= \langle a_i(x) \rangle = \langle r(x-x_i) \exp(j\psi_i) \rangle \\ &= r(x-x_i)\, M_i'(1) \\ &= r(x-x_i) \exp(j\bar{\psi}_i)\, M_i(1)\ , \end{aligned} \tag{6}$$

where M_i' is the characteristic function of ψ_i and M_i is the characteristic function of the unbiased values of phase $\psi_i - \langle\psi_i\rangle$. The last line of Eq. (6) has been written to identify the magnitude and phase components.

We consider two specific probability distributions for phase ψ_i. For gaussian distributed phase of standard deviation σ_i

$$M_i'(1) = \exp\left(-\tfrac{1}{2}\sigma_i^2\right), \tag{7}$$

and for uniformly distributed phase with total spread

$$\nu_i = \sqrt{12}\,\sigma_i\ , \tag{8}$$

then

$$M_i(1) = \operatorname{sinc}\left[\frac{\nu_i}{2\pi}\right]. \tag{9}$$

For either distribution the average amplitude transmittance at each pixel i can be controlled explicitly by selection of the value of the standard deviation σ_i. In digital simulations we prefer the uniform distribution since most gaussian random number generators are derived by performing additional numerical operations on uniform samples. A further advantage of using uniform statistics is that the total spread never need exceed 2π, whereas with gaussian statistics the standard deviation can be infinite, and which can lead to overflow and underflow errors if not handled carefully.

The expected value of Eq. (1), using either probability distribution, is the expected SLM transmittance

$$\bar{t}(x) = \sum_{i=1}^{N} \langle a_i(x) \rangle = \sum_i r(x-x_i)\, M_i(1)\, \exp(j\bar{\psi}_i) \,. \tag{10}$$

Note that because the expectation and Fourier transform operators are both linear the expected far-field transmittance (more precisely, the angular spectrum) is

$$\bar{T}(f_x) = \mathcal{F}\{\bar{t}(x)\} \,. \tag{11}$$

3.3 Summary of the design procedure

The above analysis indicates that a phase-only source distribution can be treated as a full-complex distribution for the purpose of approximating a desired far-field pattern. The results from above can be used in a design procedure as follows:

1. Specify the desired far-field pattern $T(f_x)$ or frequency domain filter.
2. For each pixel i:
   Invert Eq. (9) (preferred, or Eq. (7)) to find σ_i.
   Select a pseudorandom number, appropriately scaled by σ_i to represent the random phase ψ_i and add to the existing pixel phase.

4. HC FILTER DESIGN

The Hybrid Composite (HC) filter design[13] is a family of filters. Each HC filter in this family is a combination of four different designs; Synthetic Discriminant Function (SDF)[14], SDF Linear Phase Response (SDFLPR),[15] Minimum Average Condition Energy (MACE),[16] and Linear Phase Coefficient Composite (LPCC)[15] filters. These filters are combined into a unified form by two different parameters α_1 and α_2. Various choices of (α_1, α_2) lead to one of the above filters. Hassebrook et al., developed this filter family in an effort to understand the underlying differences and similarities between the different filter designs. They found the HC filter detection and discrimination performance was the same or superior to any one of its component filter types. We give brief background explanation and design equations for the HC composite filter.

4.1 SDF and SDF LPR

The SDF design[14] is based on linearly weighting the training images. If the $d \times N$ matrix $\mathbf{X}_t$ represents the target training image matrix, with N training images and d pixels for each image in the training-set then the SDF filter design is

$$\mathbf{h}_{SDF} = \mathbf{X}_t \mathbf{a} \,, \tag{12}$$

where $N \times 1$ vector $\mathbf{a}$ is the weight vector. In SDF design the weights are chosen to satisfy a specified response constraint at the origin of the correlation plane. If the $N \times N$ matrix

$$\mathbf{R}_{tt} = \mathbf{X}_t^T \mathbf{X}_t \,, \tag{13}$$

represents the correlation matrix for the target class, the $\mathbf{h}_{SDF}$ design is given by

$$\mathbf{h}_{SDF} = \mathbf{X}_t \mathbf{R}_{tt}^{-1} \mathbf{y} \,, \tag{14}$$

where $N \times 1$ vector $\mathbf{y}$ is the specified response of SDF at origin for the target training set. By constraining the output response to be complex with constant magnitude and linear phase, we define the SDF Linear Phase Response (SDF LPR) filters. Therefor the vector $\mathbf{y}$ for SDF-LPR is $\mathbf{y} = \lambda_n \boldsymbol{\phi}_k^*$, the magnitude value is real and positive and vector $\boldsymbol{\phi}_k^*$ is Fourier vector with element values defined by $\exp(-2\pi kn/N)$ for $k = 0, 1, \ldots, N-1$ and $n = 0, 1, \ldots, N-1$.

4.2 MACE filters

The MACE filter[16] was designed to simultaneously constrain the correlation origin response while minimizing the average correlation energy for all the training images. The resulting correlation tends to have a relatively sharp peak at origin and low sidelobes. The MACE filter was originally designed in the frequency domain and its filter equation is

$$\hat{h}_{MACE} = \hat{D}^{-1}\hat{X}(\hat{X}^{H}\hat{D}^{-1}\hat{X})^{-1}\hat{y} \tag{15}$$

where the symbol "^" represents the filter in frequency domain, "H" represents the complex conjugate of the matrix, and $d \times d$ diagonal matrix $\hat{D}$ is the energy matrix.

The spatial domain version of MACE, (SMACE),[17] is obtained by using the inverse pseudo-DFT matrix and is given by

$$h_{SMACE} = \frac{1}{d} D^{-1}X(X^{T}D^{-1}X)^{-1}y \, . \tag{16}$$

4.3 LPCC filter

LPCC filter design[13] is a family of filters. It uses phasor relationships, in conjunction with proper training set selection to obtain inherently discriminating filters. The filter equation for the kth order LPCC filter is

$$h_{LPCC} = X_t \phi_k^{*} \, , \tag{17}$$

with ϕ_k^{*} defined as before and for $k = 0,1,\ldots,N-1$. The LPCC filters are orthogonal and are combined into a filter bank for better performance.

4.4 Filter unification

The HC filter is designed to unify the aforementioned types into a single filter family. Two parameters are used for the unification. One parameter is α_1 which is used to control the variation between SMACE and classical SDF designs by weighting the sum of the spatial energy matrix D, used in SMACE, and the white noise (zero mean and variance of $\sigma^2/2$) assumed for SDF design. The resulting filter equation is

$$h_c = \frac{1}{d} P_c^{-1} X_t \left(X_t^{T} P_c^{-1} X_t\right)^{-1} y \, , \tag{18}$$

where

$$P_c = \alpha_1 D + (1-\alpha_1) N \sigma^2 I_d \, , \tag{19}$$

where the $d \times d$ matrix I_d is an identity matrix and the $d \times d$ matrix D contains the spectrum of the training set. Each element $y[n]$ of the $N \times 1$ vector y represents the origin response for the nth training image. The parameter α_1, $(0 \le \alpha_1 \le 1)$ allows incremental variation between D (i.e. training set spectrum information) and I_d (i.e., the white noise spectrum). When $\alpha_1 = 0$ we have

$$h_c = \frac{1}{d} X_t R_{tt}^{-1} y \, , \tag{20}$$

which is the classical SDF. When $\alpha_1 = 1$ we have SMACE equation.

The second parameter, α_2, is used in a matrix called "transitional smoothing matrix" and is given by

$$\underline{S}_{\alpha} = \alpha_2\left(\underline{X}_t^T \underline{P}_{\alpha}^{-1} \underline{X}_t\right)^{-1} \underline{R}_{tt} + (1-\alpha_2) N\sigma^2 \underline{I}_N \,. \tag{21}$$

The HC family of filter vectors are given by the columns of the $d \times N$ matrix, $\underline{H}_{\alpha}$ by

$$\underline{H}_{\alpha} = \underline{P}_{\alpha}^{-1} \underline{X}_t \underline{S}_{\alpha} \underline{\Phi} \,. \tag{22}$$

By varying α_1 and α_2, we can find optimum weights in the design of HC filters to yield a filter combination whose performance is better than each of the individual filters comprising the unification.

We selected the filter vectors from Eq. (22) based on their discrimination and detection capabilities.[15] The filter bank architecture[13,15] is a weighted linear sum of the magnitude responses of the selected filters. The sum is thresholded to decide whether a target of a nontarget is present.

5. PHASE-ONLY IMPLEMENTATION OF HC FILTERS USING PSEUDORANDOM NUMBER STRINGS

In using a 4f (i.e. as having input plane SLM, filter plane SLM and correlation plane CCD array) phase-only correlator we are required to encode the input images, as well as the complex filters, into phase-only format. We approach this design by first designing a filter based on phase-only input images. The resulting HC frequency domain filter is fully complex having variable amplitude and phase at each element location. There are a varicty of ways to encode the input images as well as the filters. We chose an intensity-invariant method to encode the input training images. Test images are encoded in the same way. The frequency domain filter is encoded using the pseudorandom string technique. Motivation and design methods are presented in the following subsections.

5.1 Phase-only encoding of training and test images

Images are typically acquisitioned in grey levels of intensity. Due to illumination and reflectance variations, a target image may or may not be brighter than the background and non-target imagery. We denote this variation as intensity-variance. We desire an encoding scheme that satisfies the phase-only constraint of the input plane SLM as well as being intensity-invariant.

Consider a numerically simple mapping of intensity level directly to a phase angle of a unit magnitude complex element value. For the mth element of a grey level input image vector $\underline{f}$ we have

$$z(m) = \exp\{jK f(m)\} \,, \tag{23}$$

where $Z(m)$ is the phase-only mapping and K is chosen such that

$$K = \frac{2\pi}{\max\{f(m)\}} \,. \tag{24}$$

Given this mapping, the phase is proportional to the image grey level. However, if the image varied in intensity then we could get a significantly different phase mapping, thus the mapping procedure is intensity-variant. A simple method for reducing this effect is to take the logarithm of the input image gray levels. In this way, a variation in illumination or reflectance coefficients becomes a constant phase offset across the image of interest. For example, consider an image $a\underline{f}$ where a is a reflectance coefficient and $\underline{f}$ is the image. We would have

$$z(m) = \exp\{jK(\log_2 a + \log_2 f(m))\} = \exp\{jK \log_2 a\} \exp\{jK \log_2 f(m)\} \,. \tag{25}$$

Note that the elements of $\underline{f}$ are offset such that $\min\{f(m)\} = 1$ such that the $\log_2 f(m)$ is never negative. When $\underline{z}$ is correlated with a complex filter image, a phase cancellation will occur independent of the coefficient "a." That is the magnitude of the correlation will not be affected by the value of "a." Thus we have an intensity-invariant phase-only encoding scheme. Test images may be encoded in the same manner.

5.2 Pseudorandom string encoding of frequency domain HC filters

To create an ideal HC filter family we select training images encoded to be phase-only and intensity-invariant as described in the subsection 5.1. The resulting filter family described in Eq. (22) is in the spatial domain. We perform a 2-D DFT on these filter images to obtain the frequency domain filters. These ideal frequency domain filters are encoded to phase-only filters based on the pseudorandom number algorithm described in subsection 3.3. There are conditions related to HC filter design, which require the correlation matrix structure for Linear Phase Coefficient Composite filters to approximate a circulant matrix. However, if the training set is complex than the correlation matrix $\boldsymbol{R}_{tt} = \boldsymbol{X}_t^T \boldsymbol{X}_t^*$ is, in general, Hermitian. As it turns out, if the training images have a left-right mirror symmetry which does not change with the distortion (this is the case with in-plane rotation) then $\boldsymbol{R}_{tt}$ will be real. We assume this is the case and leave the effects of a complex Hermitian correlation matrix on filter performance for future research. It should also be noted that the response constraint included on the HC filter design should compensate for the undesirable effects of a complex valued Hermitian matrix.

6. CONCLUSIONS

We have unified a phase-only encoding algorithm with a composite filter design methodology. The phase-only encoding scheme is a pseudorandom string technique introduced by Cohn et al. and the filter design was introduced by Hassebrook et al.. The result is a system design procedure for a 4f correlator. The system detection and discrimination performance is both intensity- and distortion- invariant. Issues were raised in regards to a Hermitian correlation matrix causing suboptimal performance. This issue as well as actual performance tests are left for future research.

7. ACKNOWLEDGEMENTS

This research is partially funded by ARPA and monitored by Rome Laboratory through contract F19628-92-K0021.

8. REFERENCES

1. H. Dammann and K. Gortler, "High-efficiency multiple imaging by means of multiple phase holograms," *Optics Comm.*, 3, 5, 312-315 (1971).

2. J. Jahns, M.M. Downs, M.E. Prise, N. Streibel, S.J. Walker, "Dammann gratings for laser beam shaping," *Opt. Eng.*, 28, 12, 1267-1275 (December 1989).

3. F.B. McCormick, "Generation of large spot arrays from a single laser beam by multiple imaging with binary phase gratings," *Opt. Eng.*, 28,4, 299-304 (April 1989).

4. M.P. Dames, R.J. Dowling, P. McKee, and D. Wood, "Efficient optical elements to generate intensity weighted spot arrays: design and fabrication," *Appl. Opt.*, 30, 19, 2685-91 (1991).

5. N.C. Gallagher and B. Liu, "Method for computing kinoforms that reduces image reconstruction error," *Appl. Opt.*, 12, 10, 2328-2335 (1973).

6. R.W. Gerchberg and W.O. Saxton, "Practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik* 35, 2, 237-250 (April 1972).

7. If the complex Fourier transform pair is not known then one fast Fourier transform (FFT) would need to be computed. While such direct methods are much faster than the iterative methods, it is impractical to include electronic calculation of the FFT in real-time optoelectronic processors. Optical Fourier transform and interferometric detection with video cameras can be performed at real-time rates and might be practically employed in the latter case. For example. see R.W. Cohn, "Adaptive Real-Time Architectures for Phase-Only Correlation," *Applied Optics*, 32, 5, 718-725 (10 February 1993).

8. O. Bryngdahl and F. Wyrowski, "Digital holography -- Computer-generated Holograms," in *Progress in Optics*, E. Wolf, ed., Elsevier, Amsterdam. V. 28, 1-86 (1990) and references therein including reviews by W.J. Dallas and W-H. Lee.

9. R.W Cohn and M. Liang, "Approximation fully complex spatial modulation with pseudorandom phase only

modulation" submitted to *Applied Optics*. (1993)

10. R.W. Cohn and R.J. Nonnenkamp, "Statistical Moments of the Transmittance of Phase-Only Spatial Light Modulators," *Proc. SPIE* 1751, 289-297. (1992)

11. N. Konforti, E. Marom and S.-T. Wu, "Phase-only modulation with twisted nematic liquid-crystal spatial light modulators," *Opt. Lett.*, 13, 3, 251-253. (March 1988)

12. R.W. Cohn, "Random Phase Errors and Pseudorandom Phase Modulation of Deformable Mirror Spatial Light Modulators," *Proc. SPIE*, 1772, 360-368. (1992)

13. L. G. Hassebrook, M. Rahmati, and B. V. K. Vijaya Kumar, "Hybrid composite filter banks for distortion-invariant optical pattern recognition," *Optical Engineering*, Vol. 31, pp. 923-933, May 1992.

14. C. F. Hester and D. Casasent, "Multivariant technique for multiclass pattern recognition," *Applied Optics*, vol. 19, 1980, pp. 1758-1761.

15. L. G. Hassebrook, B. V. K. Vijaya Kumar, and L. Hostetler, "Linear phase coefficient composite filters for distortion-invariant optical pattern recognition," *Optical Engineering*, Vol. 29(9), pp. 1033-1043, 1990.

16. A. Mahalanobis, B. V. K. Vijaya Kumar, and D. Casasent, "Minimum average correlation energy filters," *Applied Optics* Vol. 26, pp. 3633-3640, Sep. 1987.

17. S. I. Sudharsanan, A. Mahalanobis, and M. K. Sundareshan, "Unified framework for the synthetic discriminant functions with reduced noise variance and sharp correlation structure," *Optical Engineering*, vol. 29, 1990, pp. 1021-1028.

OPTICS LETTERS / Vol. 21, No. 4 / February 15, 1996

Random phase encoding of composite fully complex filters

Laurence G. Hassebrook, Michael E. Lhamon, and Raymond C. Daley

Department of Electrical Engineering, University of Kentucky, Lexington, Kentucky 40506

Robert W. Cohn and Minhua Liang*

Department of Electrical Engineering, University of Louisville, Louisville, Kentucky 40292

Received July 31, 1995

The mapping of complex-valued functions onto phase-only spatial light modulators is examined. Random phase encoding effectively adds amplitude control to the phase-only filter and can be used to trade off systematic errors of the phase-only filter for random errors. This is illustrated for the problem of recognizing a three-dimensional object from arbitrary views. The complex-valued composite filters that constitute a filter bank design are encoded by phase-only and pseudorandom methods. The best recognition probabilities are achieved by blending the two methods so that only the smallest amplitudes are randomly encoded. © 1996 Optical Society of America

By properly composing a filter from several views of an object it is possible to recognize the object despite distortion, thus achieving a degree of distortion invariance. One general approach has been to form a filter bank that uses multiple composite filters. The design of linear phase coefficient composite filter banks produces a set of complex-valued filters.[1] Synthetic discriminant function[2] and minimum average correlation energy[3] filters are also complex valued. These single composite filters have since been generalized into the form of filter banks.[1,4] The hybrid composite (HC) filter bank combines the properties of all these filter banks through the selection of specific values for the two parameters α_1 and α_2.[4] This study uses a specific HC filter bank that is a balanced combination of synthetic discriminant function, minimum average correlation energy, and linear phase coefficient composite properties.[4] Therefore the results reported here are relevant to a variety of composite filters of current interest.

A significant and recognized problem limiting the use of composite filters in real-time optical correlators is that currently available filter plane spatial light modulators (SLM's) do not produce full-complex modulation.[5,6] This limitation is known to modify the peak intensities and produce false peaks through nonlinear intermodulation of the composite signals.[7] Casasent and Rozzi originally noted that peak fluctuations could dramatically change peak correlation intensities while keeping only the phase of the full complex design and that even a small degree of amplitude control greatly improves recognition.[5] Previous solutions to minimize the degradation caused by limited range SLM's required numerically intensive optimization techniques.[8,9] Faster encoding procedures are needed for those applications in which the time available for optimization is a limiting factor.

Pseudorandom encoding is a specific encoding technique for mapping full-complex filters onto phase-only SLM's.[10] It is a fast procedure because it requires only one function calculation or table look-up operation per pixel. The encoding procedure adds amplitude control to the phase-only filter[11] (POF) through the addition of phase offsets $\delta\psi_i$ that have specified statistical properties. For a uniform random distribution of spread ν_i the effective amplitude control achieved at the ith SLM pixel is known to be

$$\bar{a}_i = \langle \exp(j\,\delta\psi_i) \rangle = \mathrm{sinc}(\nu_i/2\pi), \tag{1}$$

where $\langle\,\rangle$ is the expected value operator. In the encoding procedure the value of the amplitude in Eq. (1) is set to that of the desired full-complex modulation, and then Eq. (1) is inverted to yield the spread ν_i. For each pixel a randomly generated number is scaled by the appropriate spread to produce a phase offset $\delta\psi_i$ with the appropriate statistical properties. The random phase offsets are added to the phases ψ_i of the desired full-complex modulation to complete the encoding. Inasmuch as the values of Eq. (1) range between zero and one we always assume that the full-complex modulation is normalized so that its maximum amplitude is unity. Whereas no individual pixel actually produces amplitude modulation at the SLM plane, we have shown that the resulting far-field diffraction pattern is well approximated by treating each pixel as if it produced an average amplitude modulation $\bar{a}_i$.[10]

An indicator of quality of a pseudorandom encoded filter (PRF) is the diffraction efficiency (under uniform illumination)

$$\eta \equiv \frac{1}{N} \sum_{i=1}^{N} \bar{a}_i^{\,2}, \tag{2}$$

where N is the number of pixels of the SLM. The diffraction efficiency for the PRF represents the fraction of the energy illuminating the SLM that is used to form the diffraction pattern of the full-complex filter (FCF). In fact, it is exactly the diffraction efficiency of the desired, but unachievable, full-complex filter. The remaining $1 - \eta$ fraction of the energy from the PRF is diffracted into a white-noise pattern resembling speckle. With the energy divided between desired signal and random noise it is clear that, as η increases

0146-9592/96/040272-03$6.00/0

toward unity, the encoded filter will be less noisy and will more closely approximate the full-complex filter. Our searching for optimal γ is much like Juday's optimization of gain parameter G.[12] For phase-only SLM's performance is independent of G, and the traditional POF[11] always results. There is, however, an optimal value of γ.

The traditional POF,[11] for which all amplitudes of the full-complex filter are mapped to unity, may also be viewed as another type of encoding. The POF can be considered to have unity diffraction efficiency. However, as noted in Ref. 7, the encoding also introduces systematic (rather than random) errors between the desired full-complex and realized phase-only filter.

The question considered in this study is whether the amplitude control offered by pseudorandom encoding can be used to improve performance of filter banks over that possible with the phase-only filters. We analyze this problem by designing a filter bank to recognize a specific object, encoding the complex filters to POF's and PRF's, and comparing the performance of the encoded filter banks at recognizing the object in the presence of noise, clutter, and distortion. The filters are designed for implementation on a $4f$ correlator containing a 32×32 pixel amplitude-only SLM in the input scene plane and a 64×64 pixel phase-only SLM in the filter plane.

The HC filter bank design follows identically the steps described in Ref. 4. This includes the identical choice of parameters $\alpha_1 = \alpha_2 = 0.4$. In the design presented here the goal is to identify the Space Shuttle and reject all other aircraft types. The training set consists of 36 images obtained from a three-dimensional Space Shuttle model that is viewed with an altitude angle of 60°, rotated uniformly in azimuth from 0° to 360° in 10° increments, and then projected to form 32×32 pixel silhouette (i.e., binary amplitude) images. Nontarget objects are not needed for obtaining clutter-resistant HC filters and thus were not used.[4] The training images are zero padded to 64×64 pixel images. The HC filters are derived from these training images in the form of impulse responses and then fast Fourier transformed to produce the frequency plane filters. These are the filters that are encoded by various methods.

Three nontarget aircraft have also been chosen to represent clutter objects for simulations of filter bank performance. Silhouette images of these objects are taken for the same view angles and approximate scale as the target object. Representative views of the Space Shuttle and of one clutter aircraft are shown in Fig. 1. The noise shown in the figure was added only for the specific set of tests described below. Performance of the filter banks is characterized by the minimum probability of error (MPE).[4] One achieves the MPE by setting the decision threshold to produce the least total number of false alarms and misses. In our simulations we calculate MPE empirically. First we find the value of the peak response of the filter bank for the in-class object and the maximum peak value of the 3 clutter objects for each of their 36 views. The MPE is then the minimum sum of false alarms and misses divided by 72 for all possible threshold settings. Although the filter bank design produces as many filters as training images (in this case 36), one can usually achieve adequate recognition by selecting a subset of the filters that have the largest discrimination-to-noise ratios.[4] For our simulations we calculated MPE for filter banks of 1–5 filters.

In a preliminary simulation we found that filter banks using POF's usually had fewer recognition errors than filter banks using fully encoded PRF's. This result is due to the low diffraction efficiency (only a few percent) of the FCF's, which consequently introduces too much random noise. This result led us to consider a blending of encoding procedures so that only some of the pixels are pseudorandom encoded and the rest are phase-only encoded. Because the lowest amplitudes of the FCF's produced the most systematic error for phase-only encoding, we now pseudorandom encode only those amplitudes that are below a given threshold. Currently the threshold that gives the smallest value of MPE is found empirically by repeated simulations. The amount of random encoding for a given amplitude threshold can be quantified in terms akin to those for diffraction efficiency [Eq. (1)] as

$$\eta_r = \frac{1}{N} \sum_{i=1}^{N_r} \overline{a}_i^2, \tag{3}$$

where N_r is the number of pixels below threshold that are random encoded. A relative measure of the amount of pseudorandom encoding is $\gamma = \eta_r/\eta$. A γ equal to zero corresponds to phase-only encoding, and a γ of unity corresponds to pseudorandom encoding all N pixels.

The results of the simulations of MPE for various encodings and two (related) sets of test imagery are summarized in Table 1. For each set of test imagery the table presents MPE for filter banks composed of FCF's, POF's, and PRF's. The first of the two PRF columns reports the lowest value of MPE found for all values of the encoding parameter γ. The corresponding encoding parameters ranged from 0.002 to 0.07. The second PRF column reports MPE for a single fixed value of the encoding parameter ($\gamma = 0.004$) that produces reasonably low MPE's for the two sets of test imagery used. A HC filter and its pseudorandom encoding are illustrated in Fig. 2. The left side shows the gray-scale magnitudes of the filter. The right side of the figure has been binarized to indicate the pixels (in white) that are random encoded.

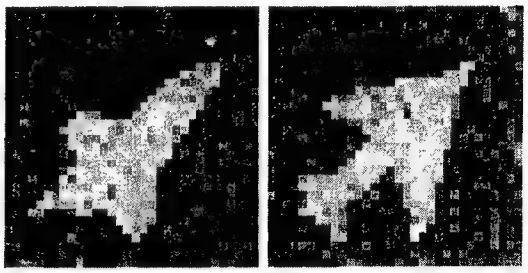


Fig. 1. One of the 36 views of the test imagery: target (Space Shuttle, left) and one of three clutter objects (Phantom, right). The noise was included in the test images only for case a of Table 1.

Table 1. Filter Bank MPE (%) for Bank Size and Type of Test Imagery

Bank Size	FCF	POF	PRF[γ]	PRF[0.004]
Case a, noisy test images				
1	22	28	27 [0.003]	28
2	7	9	6 [0.004]	6
3	6	4	2 [0.008]	3
4	1	4	1 [0.008]	1
5	4	2	0 [0.003]	1
Case b, noise-free, angle-offset test images				
1	31	40	37 [0.07]	41
2	17	28	20 [0.07]	23
3	20	25	17 [0.07]	22
4	11	24	17 [0.06]	17
5	9	23	15 [0.07]	18

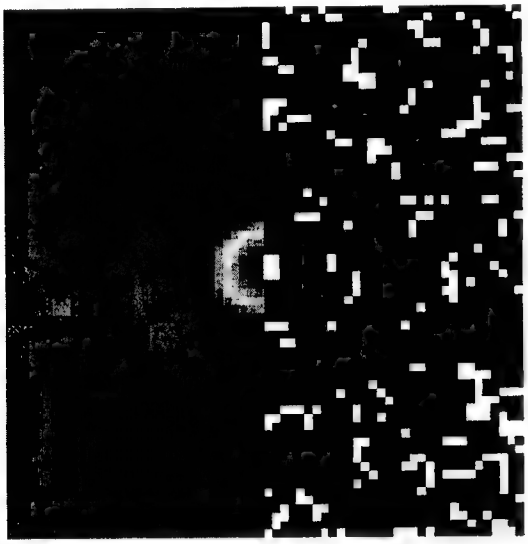


Fig. 2. Gray-level magnitudes of one HC filter. The right-hand side of filter has been binarized, indicating in white the pixels that are pseudorandom encoded for $\gamma = 0.004$.

The filter corresponds to filter number 1 in Table 1 and has a diffraction efficiency η of 4.6%. As the filter is quite nearly symmetric, the pattern of random encoding at the right of the figure will appear quite similar. For this encoding of the filter, $\gamma = 0.004$, $N_r = 480$ of the $N = 4096$ total pixels, and the maximum amplitude randomly encoded is 0.057 of the maximum (i.e., unity) filter amplitude.

To demonstrate the performance that results from different encodings we tested the filter banks against degraded imagery. In Table 1, case a, additive white Gaussian noise is added to each test image. Typical images are shown in Fig. 1. The total signal-to-noise ratio is 4:1, or 6 dB. The MPE for the 36 views is calculated 10 times, each time with a change of only the random seed for the scene noise generator. The average MPE of this test is reported in the table. In all cases the PRF has equal or lower MPE than the POF. This is true even for the case of the fixed encoding parameter $\gamma = 0.004$. However, in some cases the MPE for the PRF, and even for the POF, is lower than that for the FCF. These crossovers are not inconceivable when one recognizes that α_1 and α_2 have been selected to give best overall performance for various scene distortions, noise, and clutter and are not necessarily the optimal choices, for any one environment. Therefore phase-only encoding could, by chance, have lower MPE.

In case b the test images are distorted views of the original training and clutter objects. The views are taken with a fixed angular offset in azimuth from the original 36 views. The MPE is calculated for each offset, 1° to 9°, in 1° increments, and the average of the MPE's is presented in Table 1. With larger MPE's the filter bank learning curves appear more stable and the differences are more easily seen. For more than one filter, both PRF columns have lower MPE than the POF. The PRF with optimized MPE for two and three filters has MPE that is comparable with that for the FCF.

An additional trend noted in Table 1 is that the encoding parameter γ appears to increase as the discrimination task becomes more challenging. Case b, which has the most errors for a given number of filters, typically has the largest values of γ. This means that more of the pixels are pseudorandom encoded for this case. Further study is needed to determine whether the trend is reliable and, if it is, the reason for it.

We have shown that pseudorandom encoding the lowest amplitude pixels of composite, fully complex filters can noticeably improve recognition performance over that which is possible with the traditional phase-only encoding. Although pseudorandom encoding does not produce the optimal mapping for correlation, it does provide some level of improvement in recognition and is easy to use with real-time hardware. For the study we did searches to find the best threshold for random encoding. However, a fixed encoding threshold can often produce improved recognition.

This research was supported by NASA cooperative agreement NCCW-60 through Western Kentucky University and by Advanced Research Projects Agency contract F19628-92-0021 through Rome Laboratory.

*Present address, Neuristics Corporation, Baltimore, Maryland 21204-2316.

References

1. L. G. Hassebrook, B. V. K. Vijaya Kumar, and L. Hostetler, Opt. Eng. **29,** 1033 (1990).
2. C. F. Hester and D. Casasent, Appl. Opt. **19,** 1758 (1980).
3. A. Mahalanobis, B. V. K. Vijaya Kumar, and D. Casasent, Appl. Opt. **26,** 3633 (1987).
4. L. G. Hassebrook, M. Rahmati, and B. V. K. Vijaya Kumar, Opt. Eng. **31,** 923 (1992).
5. D. Casasent and W. A. Rozzi, Appl. Opt. **25,** 3767 (1986).
6. B. V. K. Vijaya Kumar, Appl. Opt. **31,** 4773 (1992).
7. J. A. Davis and D. M. Cottrell, Opt. Lett. **19,** 496 (1994).
8. D. A. Jared and D. J. Ennis, Appl. Opt. **28,** 232 (1989).
9. Z. Bahri and B. V. K. Vijaya Kumar, Proc. SPIE **1151,** 138 (1989).
10. R. W. Cohn and M. Liang, Appl. Opt. **33,** 4406 (1994).
11. J. L. Horner and P. D. Gianino, Appl. Opt. **23,** 812 (1984).
12. R. D. Juday, Appl. Opt. **32,** 5100 (1993).

Ref. 12 is not available at this time.

(Submitted to Journal of the Optical Society of America, A, 7 February 1996)

Fully Complex Diffractive Optics via Patterned Diffuser Arrays

Robert W. Cohn, Anatoly A. Vasiliev, Wenyao Liu* and David L. Hill
University of Louisville, Department of Electrical Engineering, Louisville, Kentucky 40292
(502) 852-7077, Fax: (502) 852-6807, Rwcohn01@Ulkyvm.Louisville.Edu
* Permanent address: Tianjin U., Dept. of Precision Instrument Engineering, Tianjin, China 300072

Abstract

The properties of random phase are applied to the synthesis of fully complex aperture functions. Arbitrary functions can be envisioned and also implemented as arrays of individually specified diffusers. For any given diffuser pixel, two parameters, average step height and vertical roughness independently control phase and amplitude. Noise introduced by roughness can be quite low for many designs of interest, especially if the roughness pattern is of much finer scale than the pixel dimensions. A potentially low cost fabrication method is proposed in which the desired pixel topography is patterned by exposing photoresist with partially developed laser speckle patterns.

Keywords. Signal synthesis, kinoforms, computer generated holography, rough surface scattering, laser speckle, micro-fabrication of three dimensional surfaces, optical information processing, statistical optics, phased arrays

1. Introduction

The properties of random phase have been widely applied to analyze the scattering of monochromatic light from rough surfaces.[1,2] The inverse problem of specifying the statistical properties of phase-only structures so as to obtain desired far-field diffraction patterns has received little attention. Recently Cohn and Liang introduced a point oriented encoding method, referred to as pseudorandom phase-only encoding, in which phase modulation drawn from a custom-specified nonstationary random process is used to represent the desired amplitude modulation.[3] While pseudorandom phase codes are widely used in optical information processing, holography and optical memory storage, these approaches all appear to only use pseudorandom phase sequences that are totally random (e.g. phase uniformly distributed over a 2π range.) Instead, the pseudorandom encoding method individually specifies phase randomness at each point in proportion to the reduction in amplitude transmittance (or reflectance) desired. We use pseudorandom encoding concepts to show how arrays of diffusers can produce nearly arbitrary diffraction patterns.

Introduction to complex modulating property of diffuser pixels. The effective amplitude modulation at a point corresponds to the specular component scattered from that point. This is most easily seen by considering the formation of partially developed speckle patterns as illustrated in fig. 1. The far-field diffraction pattern can be varied from purely specular to purely diffuse scatter (i.e. a fully developed speckle pattern.) As drawn, the roughness is of a much higher spatial frequency than the illumination footprint, and thus the specular component can be much brighter than the diffuse component for a wide range of roughnesses. For purposes of diffractive optics design the specular component represents the desired diffraction pattern and the diffuse component represents noise.

Near arbitrary diffraction patterns, only limited by speckle noise, can be produced by superposition of the specular components from arrays of surfaces of specified roughness (e.g. the surface shown in fig. 2.) This can be shown as follows: The modulation of a plane wave reflected from a phase-only surface is represented by the complex-valued (indicated by bold) function $\boldsymbol{a}(x,y)$= $\exp[j\psi(x,y)]$. The far-field diffraction pattern of the modulation pattern is $\boldsymbol{A}(f_x,f_y) = \mathcal{F}[\boldsymbol{a}]$ where $\mathcal{F}[\bullet]$ is the fourier transform operator. Since the fourier transform and ensemble average are both linear, the average complex-valued far-field pattern of a random complex pattern can be written

$$\langle \boldsymbol{A} \rangle = \mathcal{F}[\langle \boldsymbol{a} \rangle] \tag{1}$$

where $\langle \bullet \rangle$ is the expectation operator. Under the assumption that the random samples of $\boldsymbol{a}$ are statistically independent with position, the expectation of I the far-field intensity pattern is

$$\langle I \rangle = \langle |\boldsymbol{A}|^2 \rangle = |\langle \boldsymbol{A} \rangle|^2 + \langle I_s \rangle \tag{2}$$

where $I_s(f_x,f_y)$ is a residual noise pattern due to the random phasings in the far-field.[3] As long as the noise [represented by the second term of eq. (2)] is adequately low, then eq. (2) is approximately the magnitude squared of eq. (1). In this average sense, any complex-valued modulation can be represented by the random phase-only modulation $\boldsymbol{a}(x,y) = \exp[j\psi(x,y)]$ using the relationship

$$\langle \boldsymbol{a} \rangle = \int p(\psi) \exp(j\psi) d\psi = a_p \exp(j\langle \psi \rangle) \tag{3}$$

where $p(\psi)$ is the probability density function (pdf) of the phase and a_p is the resulting expected amplitude modulation. We will often refer to a_p as the *effective amplitude*, $\phi_p=\langle\psi\rangle$ as the *effective phase*, and $\boldsymbol{a}_p \equiv \langle \boldsymbol{a} \rangle$ as the *effective complex amplitude* or *modulation.*

A desired modulation $\boldsymbol{a}_c(x,y)$ is produced by specifying a probability density function $p(\psi)$ in eq. (3) that gives $\langle \boldsymbol{a} \rangle = \boldsymbol{a}_c$. The actual value of phase is selected from a random number generator having the required density function. Amplitudes can be encoded using simple pdf's; for example, the uniform family of density functions, with spreads $\nu \in [0,2\pi]$ when evaluated in eq. (3) gives all values of amplitude between 0 and 1 according to

$$a_p = \operatorname{sinc}(\nu / 2\pi) \tag{4}$$

Thus the correct density function for implementing a particular value of a_p can be found by inverting eq. (4) for the appropriate value of ν. The most widely available random number generator routine is uniform with a spread of 1 and a mean of ½. A number selected by this routine would be scaled by υ and offset by ϕ_p-ν/2 to produce the actual random phase ψ. The procedure is repeated for all points to specify the analog phase-only function $\boldsymbol{a}(x,y)$ that represents the desired modulation $\boldsymbol{a}_c(x,y)$.

The pseudorandom encoding method, as described above, can be directly applied to designing a kinoform. One could spatially sample $\boldsymbol{a}_c$ up to the resolution of the pattern generation system, and then apply eqs. (3) and (4) to each sample. Alternately, if the bandwidth of $\boldsymbol{a}_c$ is lower than half the sampling rate of the pattern generator then, according to nyquist theory, $\boldsymbol{a}_c$ can be adequately represented with fewer samples. This is desirable for reducing machine time. However, for pseudorandom encoding there is still a significant advantage if the sampling rate is high. That is, as illustrated in Fig. 1, the spatial extent of the speckle noise will be greatest, and thus its intensity will be lowest, for the highest sampling rate. It would seem that ψ_c the phase of $\boldsymbol{a}_c$ could be sampled at the lower rate and the amplitude a_c could be sampled at the highest rate with little loss. This

would result in the structure shown in fig. 2., which is an array of diffusers for which each diffuser has a unique step height and for which the vertical roughness varies (depending on the effective amplitude desired) from diffuser pixel to diffuser pixel. Fabrication of such a diffuser by writing each individual random phase value in sequence would require a high resolution and costly pattern generator. Instead we propose methods for fabricating each diffuser pixel in a single patterning step that appear to be less costly than traditional pattern generators. Before describing the fabrication system, we compare the patterned diffuser array design approach with prior kinoform design procedures and we review the prior fabrication approaches.

Relationship to prior kinoform design procedures. Currently numerically intensive global search and optimization algorithms are widely used for synthesizing modulation functions under the constraint of phase-only (in many cases binary phase-only) modulation.[4-8] Depending on pixel dimensions or the maximum spatial bandwidth of the desired complex modulation, direct pixel-by-pixel or point-by-point encoding can be a practical alternative. Several methods of encoding complex functions onto phase-only diffractive structures were developed shortly after the introduction of the kinoform.[9,10] The most direct is the Kirk and Jones method[11] where a periodic carrier of spatial frequency f_0 that is modulated in amplitude α and phase ψ_α is converted into the phase-only function

$$\boldsymbol{a}(x,y) \;=\; \exp\left[j\,\psi(x,y)\right] \;=\; \exp\left\{j\left[\alpha\, h(2\pi f_0 x) + \psi_\alpha(x,y)\right]\right\} \quad . \tag{5}$$

One specific case considered by Kirk and Jones was for $h(\bullet) = \cos(\bullet)$. For this case the Fourier series expansion of eq. (5) produces a D.C. component of complex amplitude

$$\boldsymbol{a}_c \;\equiv\; a_c \exp\left(j\psi_c\right) \;=\; J_0(\alpha)\exp\left(j\psi_\alpha\right) \tag{6}$$

where $J_0(\alpha)$ is the zero order Bessel function. Thus $\boldsymbol{a}_c$ is proportional to the complex amplitude of the D.C. or zero order far-field diffraction pattern. Any desired value of amplitude a_c between 1 and 0 can be implemented by inverting $J_0(\alpha)$ to find the appropriate value of α. Similar results can be developed for $h(\bullet)$ a square wave carrier, and also for a rectangular carrier of variable duty cycle.

From the perspective of the Kirk and Jones approach, patterned diffuser arrays use a random carrier. That is to say, rather than use a single frequency carrier $h(\bullet)$, the carrier is a randomly phased combination of a continuous range of frequencies. Whereas, the traditional Kirk and Jones method scatters unwanted energy into a set of harmonics at discrete frequencies, a random carrier scatters unwanted energy uniformly (on average) into all frequencies of the carrier. In many applications a low level of diffuse background noise in-band, may be preferable to the presence of harmonics out of band. As far as fabrication, it may be easier to synthesize the random carrier than a single frequency carrier. For any of these carrier based methods it is important to note that the maximum useful diffraction efficiency of $\boldsymbol{a}(x,y)$ is only limited by the efficiency of the desired complex modulation $\boldsymbol{a}_c(x,y)$. Thus, there is no implementation loss for the on-axis diffraction order. Furthermore, the optimization of the function $\boldsymbol{a}_c$ required to meet a specific set of design criteria, is decoupled from the constraints imposed by the phase-only implementation. This could potentially lead to simplified and improved diffractive optic design procedures. (For instance non-iterative optimal window design procedures become possible; see ref. 3 for a specific design of a top hat far-field pattern.)

Comparison with prior fabrication methods. Kirk and Jones also presented a fabrication

procedure in which a photomask having a sinusoidally varying intensity transmittance is placed in contact with a photographic recording medium for which thickness depends linearly on exposure energy. The medium is exposed with an intensity pattern proportional to the function $\alpha(x,y)$. Then the mask is removed and the medium is further exposed with a second pattern proportional to $\psi_\alpha(x,y) = \psi_c(x,y) + 2\pi - \alpha(x,y)$ that adjusts thickness to produce the desired phase modulation $\psi_\delta(x,y)$. [The term $2\pi - \alpha$ compensates for the average thickness variations introduced by $\alpha\, h(\bullet)$.]

If a square wave carrier is used instead of a sinusoid, the photomask is much easier to produce. If a rectangular carrier is used, the duty cycle is varied. This has the advantage that every pixel can be exposed with the same dose, but it has the disadvantage that the photomask must be written with extreme precision and a custom photomask is needed for each new device design. Also, all three deterministic carriers (sinusoidal, square and rectangular) require two exposures to produce a desired complex value at a point. A single exposure method can be envisioned in which laser interference is used to produce sinusoidal fringes, and beam balance is adjusted to control phase bias. This pattern would be projected through a small aperture and the entire photographic medium would be exposed by translating it under the aperture. This method, of course, requires good fringe stability.

The Kirk and Jones approach does not seem to have been widely used; apparently, because of the requirement for analog control of the exposure. Currently it is most common to fabricate computer generated diffractive optical elements as binary and m-ary phase steps. However, lately there has been considerable progress in producing analog phase-only relief structures. Various approaches include projection printing, and laser or electron beam direct write onto photoresist.[12-16] For highly resolved structures current laser beam scanning can be quite slow.[16] E-beam scanning can be fast if there are a limited number of mechanical steps between fields, but e-beams are generally very expensive to purchase and maintain. Both e-beam and laser systems have very accurate, precise and expensive positioning systems. As a result of these recent developments in fabrication art we propose an alternate approach to topographic patterning of photoresists. In particular, we consider the possibility of producing patterned diffusers arrays and the technical issues that would affect the quality of the resulting diffraction patterns.

2. Advantage of Diffuser Pixels over Single Step Pixels: Directionality Gain

A random rough surface or diffuser can be modeled as an array of random step heights. According to the law of large numbers,[17] increasing the number of random steps, or equivalently phase values, across the diffuser will make the far-field diffraction pattern more predictable. For a partially developed speckle pattern the specular component will be more clearly seen over the noise for diffusers having a larger number of phase samples. This effect can be interpreted as a directionality gain of the specular component over the diffuse component. If there are N statistically independent roughness samples, or cells, filling an aperture then the intensity of the far-field diffraction pattern will be reduced by a factor of $1/N$ over there being one roughness cell filling the aperture. Since the diffraction pattern of the single roughness cell is identical to the pattern of the uniformly illuminated aperture, then the directionality gain of specular to diffuse is N.

Fig. 3 illustrates the improvement resulting from increasing the number of rough samples from 1 to 9 per pixel for a diffractive optical element (DOE) designed to produce an 8x8 array of uniform intensity spots. Both DOEs represent the same 100x100 fully complex array of numbers whose

amplitudes are encoded into phase according to eq. (4). Rather than performing the encoding step once per pixel (as originally prescribed by ref. 3) it is performed 9 times for the 3x3 array of phases used to represent each diffuser pixel. The phases of the i'th pixel are randomly selected from the uniform random distribution having spread v_i. The top row of fig. 3 shows the resulting diffraction patterns as simulated using the fast Fourier transform (FFT) and the bottom row shows the result for diffraction from a Hughes birefringent liquid crystal light valve that is programmed to approximate the desired phase modulation. As anticipated, the photographs show that the speckle is more broadly scattered and its intensity is reduced by using diffuser arrays. The non-uniformity of the diffraction pattern (defined as the standard deviation of the intensity of the 64 spots divided by average intensity of the 64 spots) improved from 23.6 % to 9.9 % by using diffusers in place of single phase pixels. The signal to noise ratio (defined as the average intensity of the 64 spots to average background intensity) also improved from 72 to 1639. For the experimental intensity patterns the as-measured non-uniformity improves from 36.9 % to 30.4 % . However, the images indicate that the intensity gradually decreases with distance from the optical axis. This is due to limited resolution of the SLM (which includes rolloff in the video output of the frame grabber and the cathode ray tube that is the write light source for the light valve.) Applying a least squares regression (of constant, linear and quadratic factors in both x and y) we have removed this systematic trend. With the trend removed the nonuniformity due to statistical fluctuation alone is 29.4 % using single pixels and 21.6 % using diffuser pixels. (For comparison the nonuniformity for the theoretical images after removing linear and quadratic trends was only reduced about 2 % in each case.) For the experimental measurements, the signal to noise ratio (in the vicinity of the spot array) also increased from 14 to 35 by using the diffuser pixels. Even though the experimental spot arrays are less uniform and more noisy than theory (due to loss of resolution and inexact phase control of the SLM) the improvements possible using diffuser pixels are apparent. Furthermore, the experiment shows that applying pseudorandom encoding to non-ideal devices produces results that are qualitatively similar to theory.

3. Microtopographic Patterning Methods

Our goal is to develop a robust, repeatable and easy to implement patterning technique. While, in concept, we can write one random phase at a time by direct pseudorandom encoding [eqs. (3-4)] there is really no need for this precise and detailed control. Instead we can directly use the statistical properties of laser speckle which is known to be reproducible and controllable.

Fig. 4a illustrates one basic pattern generator concept. This apparatus is a type of proximity printer. An aperture (perhaps patterned on a chrome photomask) having the area of a diffuser pixel is kinematically supported as close to the photoresist as practical. The photoresist is exposed through the aperture and then the substrate is translated to the next location to be exposed. The high spatial frequency random carrier is a fully developed speckle pattern generated by the ground glass diffuser. An average intensity offset needed to produce a phase bias can be generated by temporal averaging of speckle patterns. This can be achieved, as illustrated in fig. 4a, by spinning a ground glass diffuser with a constant angular velocity. The radial separation between the beam and the diffuser axis determines the linear velocity of the diffuser. Linear velocity together with exposure time then determines the effective bias. A theory for this is described in Sec. 4.

Fig. 4b shows a modified approach in which a uniform intensity pattern can also be used to provide a phase bias. Statistical properties of the intensities of coherently biased speckle patterns

are described in ref. 1. We specifically consider the case in which the bias and speckle pattern are mutually incoherent. For the second approach, the uniform and speckle illumination could obviously be combined with a beam splitter. However, in order to eliminate beam splitter loss and multiple reflections it is possible to bring a uniform coherent illumination through a small aperture (say a fiber optic) in the diffuser, as illustrated in fig. 4b. Mutual coherence between the spatially uniform and non-uniform sources can be achieved by rotating a polarized fiber into the cross polarized state or by using the fiber to introduce a delay difference in excess of the coherence length of the laser.

A third approach would be to simply apply appropriate random signals to the exposure control signal on an electron beam or laser beam direct write system. The only advantages of this technique over previous direct written DOEs is that the complexity of the design procedure is simplified and the number of values placed in machine memory can be greatly reduced.

We briefly mention two other potential applications of the diffuser array concept. Liquid crystal under applied voltage can be converted between isotropic and randomly oriented states. It may be possible to develop a real-time spatial light modulator in which this type of liquid crystal layer is cascaded with pure phase retarding pixels. We present this device more to illustrate the concept of diffuser arrays than as a serious candidate device. The currently prevailing view is that the development of any tandem SLM is viewed as too costly and risky. The second application is to use patterned diffuser arrays as grayscale masks in projection printers. These masks could be used in place of true gray scale and halftone masks that were recently used to demonstrate projection printing of three dimensional diffractive optical structures in photoresist.[13,14] For either the half-tone mask or the pseudorandom patterns grayscale is achieved by diffracting light outside the aperture of the imaging lens. Speckle would not be present in the projected image if the source illumination is adequately incoherent. The grayscale effect can be easily demonstrated by placing a piece of ground glass on the platen of an overhead projector. The pseudorandom masks for projection printing could be fabricated with either system shown in fig. 4. The remainder of this paper considers technical issues associated with the patterning systems shown in fig. 4.

4. Technical Considerations for Patterning Diffuser Pixels in Photoresist

Issue 1: Proximity recording of laser speckle. Projecting laser speckle through a small aperture may unacceptably blur the exposure pattern. As an example consider fig. 5 which shows how a fully developed speckle pattern (457 nm argon-ion wave length) diffracts at various distances past a 100 μm slit. At a distance of 100 μm past the slit the edges of the pattern are still rather sharp. This indicates that pixels having a large fill factor can be made by proximity exposure for reasonable separations between the mask and resist. For a uniform bias illumination (coherent or incoherent) and a separation of 100 μm the transition from light to dark (due to Fresnel diffraction) is approximately 5 μm.

As compared to recording interference fringes, speckle requires minimal vibration isolation. For a diffuser, laser and ccd observation camera on a 2" thick optical breadboard supported by a wood table we observed that speckle patterns displayed on a video monitor exhibited no apparent displacement for speckle diameters larger than 2 μm. Vibration was noticeable for .6 μm speckle but no blurring was observed for 1/30 second exposure frames recorded using a frame grabber. Thus it seems that it is quite practical to illuminate resist with 2 μm speckle though an aperture in near

contact (100 μm or less). For pixels on the order of size of current SLM pixels (12.5 - 100 μm) the directivity gains can be 39 to 2500.

Issue 2: Complex Modulation for Recording Speckle in Linear Resist. The probability density function of I_s the intensity of fully developed speckle is known to be exponentially distributed[1,2] and is written

$$p(I_s) = \frac{1}{\langle I_s \rangle} \exp\left[\frac{-I_s}{\langle I_s \rangle}\right] \tag{7}$$

where $\langle I_s \rangle$ is the average intensity of the speckle pattern. Also, since speckle intensity is exponentially distributed, $\langle I_s \rangle$ can be interpreted as the standard deviation of the speckle intensity. For a photoresist which linearly maps exposure energy into resist thickness, ψ_s the random phase depth produced is proportional to exposure energy E_s and intensity I_s of the speckle pattern. Likewise, a mutually incoherent and spatially uniform illumination can be used to produce a bias phase shift ψ_b so that the total phase random phase shift can be expressed $\psi = \psi_b + \psi_s$ where ψ_b is proportional to E_b, the bias exposure. The effective complex modulation produced by this surface can be found by treating the actual phase depth ψ as an exponentially distributed random variable. Using the pdf for ψ of the form of eq. (7) in eq. (3) yields

$$\langle a \rangle = \frac{\exp\left[j\left(\psi_b + \operatorname{atan}\langle \psi_s \rangle \right) \right]}{\sqrt{1 + \langle \psi_s \rangle^2}} \quad . \tag{8}$$

The amplitude decreases monotonically with increasing average phase depth of the resist $\langle \psi_s \rangle$ (which is also proportional to average energy density of the speckle $\langle E_s \rangle$.) The phase shift due to speckle alone only varies from zero to $\pi/2$, but ψ_b can be chosen to produce any phase shift from zero to 2π.

Issue 3: Selecting Resist Thickness to Ensure Linearity. A linear resist will effectively saturate if developed through its entire thickness down to the substrate. This nonlinearity will change the complex modulation over that predicted by eq. (8). Consider that the total resist thickness is proportional to the maximum phase shift $\psi_m = \psi_b + \psi_{ms}$ where ψ_{ms} is the maximum phase shift available for speckle recording at a given bias. The effective complex modulation for this case is found by evaluating eq. (3) as

$$\langle a \rangle = \exp(j\psi_b) \times \left[\int_0^{\psi_{ms}} p(\psi) \exp(j\psi) d\psi + \exp(j\psi_{ms}) \int_{\psi_{ms}}^{\infty} p(\psi) d\psi \right] \quad . \tag{9}$$

where the density function is of the exponential form in eq. (7). This evaluates to

$$\langle a \rangle' = \frac{\exp\left[j\left(\psi_b + \operatorname{atan}\langle \psi_s \rangle \right) \right]}{\sqrt{1 + \langle \psi_s \rangle^2}} \left\{ 1 + \langle \psi_s \rangle \exp\left[j\left(\psi_{ms} - \pi/2 \right) \right] \exp\left(\frac{-\psi_{ms}}{\langle \psi_s \rangle} \right) \right\} \quad . \tag{10}$$

where the prime is used to indicate that this result is perturbed from the result in eq. (8). If the saturated value of phase ψ_{ms} is much greater than $\langle \psi_s \rangle$ the average phase produced by a purely linear recording of speckle then eq. (10) reduces to eq. (8). Thus the second term in braces of eq. (10)

represents the errors due to finite resist thickness. A minimum thickness can be selected based on the minimum amplitude a_{min} of $a_c \in [a_{min}, 1]$ that is practical to implement, and the maximum allowable error ε between eq. (10) and eq. (8). The worst case absolute error is approximately

$$\varepsilon(a_c) \approx \exp(-a_c \psi_{ms}) \tag{11}$$

where the approximation $a_c = |\langle a \rangle| \approx 1/\langle \psi_s \rangle$ for average phase depth much greater than one radian has been used in eq. (10). The minimum total resist thickness is then proportional to

$$\psi_t \equiv \psi_{mb} + \psi_{ms} = 2\pi - \ln(\varepsilon_{min})/a_{min} \tag{12}$$

where $\varepsilon_{min} = \varepsilon(a_{min})$ and $\psi_{mb} = 2\pi$ is the maximum bias shift required to achieve all possible phase shifts. Using ψ_{ms} as defined in eq. (12) in eq. (11) gives error as a function of a_c of

$$\varepsilon(a_c) = (\varepsilon_{min})^{a_c/a_{min}} . \tag{13}$$

As a specific example of using these equations to select resist thickness consider the case for a minimum amplitude of a_{min}=.025 and an absolute error ε_{min}=.0025, or 10 % relative error. Using eq. (12) the resist thickness is ψ_t = 246 radians or 39.1 optical wavelengths. For a reflective surface relief pattern, and an optical wavelength of .633 μm, the resist can be as thin as 12.3 μm. Eq. (13) shows that the relative error decreases rapidly for a_c greater than .025 . For example, for a_c=.03 the error drops to .00075 . Resist thickness is then only a significant concern for very small amplitudes, i.e. those values smaller than .025 . The thickness is quite reasonable for standard photoresists.[18]

For comparison, the Kirk and Jones method using a sinusoidal carrier requires a thickness of at least

$$\psi_t = 2\pi + 2 J_0^{-1}(a_{min}) \tag{14}$$

which follows from eqs. (5) and (6). For a_{min} equal zero the total thickness for a reflective surface is .56 μm. While the thickness of the resist using for the random method is much larger than the deterministic method, it should be recognized that the selection of thickness in eqs. (11) and (12) used a worst case design. Furthermore, the maximum average speckle exposure energy is proportional to $\langle \psi_s \rangle \approx 1/a_{min}$ which corresponds to an average depth of 2 μm. Thus the comparison in terms of energy use is more favorable. The basic conclusion for these example numbers is that the resist can be treated as infinitely thick for resists 6 times more thick than the average speckle depth.

The pseudorandom method can also produce an effective zero. If the magnitude of the second term in the braces in eq. (10) is unity and $\psi_{ms} = \psi_t - \psi_b$ and $\langle \psi_s \rangle$ are chosen to produce a phase shift of π, then eq. (10) is zero. This is equivalent to having a relative error of 100 % between eqs. (8) and (10). For example, for the 39.1 wavelength thick resist discussed above, an exposure depth of 9.6 wavelengths or 3.05 μm produces a zero according to eq. (10) as compared to an amplitude of a_c = .0165 for an infinitely thick resist [according to eq. (8).] Unless the exposure system is precisely controlled and the resist thickness is precisely known it would actually be quite difficult to accurately implement a true zero by this method. In most applications a very low minimum effective amplitude should be adequate.

Issue 4: Transformation of Speckle Statistics by Recording in Log Nonlinear Resist. For many resists, thickness is proportional to the logarithm of exposure over a wide dynamic range. For such resists the exposure curve (depth into the resist t vs. exposure energy E) takes the form

$$t(E) = m \ln\left(E/E_b\right) \tag{15}$$

where E_b is a reference recording level corresponding to a reference thickness t=0 and m is the logarithmic slope of the resist. The exposure curve of a 9.5 μm thick film of resist (AZ 4903 positive) presented in ref. 15 is well fit over a 7 μm range for a slope of m=2.70 μm and a reference energy E_b=75 mJ/cm^2. For 5 μm films of Shipley S1650 resist we have experimentally determined that the slope is m=.823 μm over a 2.6 μm range starting from a reference energy of E_b=40 mJ/cm^2.

Using the logarithmic range of a resist leads to effective amplitude that depends on the ratio of the speckle exposure to bias exposure rather than absolute intensity. This may prove to be an advantageous feature since it is often easier to control ratios (using a halfwave plate and a polarized beamsplitter) than it is to individually control the absolute energy in two independent exposures.

The effective amplitude can be found using the following analysis. The logarithmic recording medium produces the total phase shift

$$\psi_t = \psi_b + \psi_s = \alpha \ln\left(E_s + E_b\right) \tag{16}$$

where α is the logarithmic slope in radians (i.e. $\alpha = 4\pi m/\lambda$ for a reflective surface) and $\psi_b = \alpha \ln(E_b)$. This definition allows the phase shift due to speckle to be written as

$$\psi_s = \alpha \ln\left(1 + E_s/E_b\right) \tag{17}$$

Using the definitions in eqs. (15) and (16), the exponential density of the form of eq. (7) and the change of variables $x = E_s/\langle E_s\rangle$ in eq. (3) leads to

$$\begin{aligned} \langle a \rangle &= \exp\left(j\psi_b\right) \int_0^\infty \exp(-x)\, \exp\left[j\alpha \ln(1+\gamma x)\right]\, dx \\ &= \exp\left[j\left(\psi_b + \alpha \ln\gamma\right)\right] \exp(1/\gamma)\, \Gamma(1+j\alpha, 1/\gamma) \end{aligned} \tag{18}$$

where $\gamma = \langle E_s\rangle/E_b$ and Γ(a,b) is the incomplete gamma function.[19] Fig. 6 shows the effective amplitude produced by exposing the S1650 and AZ4903 resists (described above) with speckle patterns and then reflecting 633 nm light from the resulting surfaces. This corresponds to using α=16.34 and 53.6 in the evaluation of eq. (18). For these values of α the effective phase (excluding bias ψ_b) varies by slightly more than $\pi/2$ for all values of γ. This amount of phase modulation is comparable to the maximum phase shift for linear resists [see eq. (8) and fig. 6].

The minimum resist thickness that effectively behaves as infinitely thick [thus permitting the use of eq. (18)] can be determined by an analysis similar to that in *Issue 3*. The maximum phase shift for which the resist is exposed down to the substrate is once again written as $\psi_m = \psi_b + \psi_s$. However, the maximum phase shift due to speckle is now explicitly written as $\psi_{ms} = \alpha \ln(1 + E_{ms}/E_b)$ where E_{ms} is the amount of energy above bias at which the resist is completely exposed. With these definitions the perturbed version of eq. (18) is written

$$\langle a\rangle' = \langle a\rangle + \varepsilon = \langle a\rangle + \exp\left(j\psi_b\right)\left\{\exp\left(\frac{-\gamma_{ms}}{\gamma} + j\psi_{ms}\right) - \int_{\gamma_{ms}/\gamma}^{\infty} \exp\left[-x + j\alpha\ln(1+\gamma x)\right] dx\right\} \qquad (19)$$

where the definition $\gamma_{ms}=E_{ms}/E_b$ has been used and ε is the absolute error resulting from the perturbation.

Continuing with the numerical example begun in *Issue 3*, a value of γ is found using eq. (18) for which a_c=.025 . For the resist with the smaller logarithmic slope (α=16.34) a value of γ of 2.45 is needed to produce this amplitude. For the resist with the larger slope a value of γ= .745 is needed. For an absolute error ε of less than .0025 then the ratio $\gamma_{ms}/\gamma=E_{ms}/\langle E_s\rangle$ needs to be approximately 6 or greater [as found by numerical evaluation of eq. (19).] This is essentially identical to the result for linear photoresist. However due to the nature of the logarithmic resists the resist thickness can be much less than for linear resists. The minimum thicknesses are 2.26 μm for the low α resist, 4.59 μm for the high α resist as compared to 12.1 μm for the linear resist. The required thickness can be appreciated by comparing it against the probability density function for the recorded depths (which are proportional to the random phases ψ_s .) This is shown in fig. 7. The density function for logarithmically recorded speckle has been derived by a standard technique for transformations of random variables.[17] This function is written

$$p(\psi_s) = \frac{1}{\alpha\gamma}\exp\left\{\frac{\psi_s}{\alpha} + \frac{1}{\gamma}\left[1 - \exp\left(\frac{\psi_s}{\gamma}\right)\right]\right\} \qquad (20)$$

Note that for each curve in Fig. 7 $p(0)$=.025= a_c . Also note that for the logarithmic resists $p(0)=1/(\alpha\gamma)$. The relationship between the pdf and effective amplitude is approximately valid for α greater than or equal to 4. For α near 5 the pdf curve is even more sharply peaked and narrow than for the α=16.34 curve and the maximum resist thickness is around 1 μm. For α greater than 53.6 the pdf more closely approaches the exponential distribution for a linear resist. For an appropriately chosen value of α, logarithmic transformation of speckle permits the use of much thinner films than for linear resists.

Issue 5: Special case: low sensitivity log resist. For resists having sensitivities below 4 the effective amplitude cannot be continuously controlled between one and zero. This can be seen by evaluating eq. (18). For large values of γ the effective amplitude is well approximated as

$$\langle a\rangle \approx \exp\left[j\left(\psi_b + \alpha\ln\gamma\right)\right]\Gamma(1+j\alpha) \qquad (18a)$$

where $\Gamma(\cdot)\equiv\Gamma(\cdot,0)$ is the gamma function. The magnitude of eq. (18a) decreases monotonically with increasing α. For example for a_p=.01, .25, .5, .75 then α=4, 1.62, 1.04, .625 respectively. The effective amplitude as a function of γ [as calculated using eq. (18)] can oscillate around the limiting value of effective amplitude, but this is usually a negligible amount. The only significant undershoot is evident for α close to e=2.72 . In this instance the effective amplitude as a function of γ dips to zero (at $\gamma\approx5$) before settling to an effective amplitude of .058 . The most important point is that a high contrast material (α>4) is required in order to produce fully complex modulation.

Issue 6: Time-average recording in linear resists. Consider recording M equal intensity,

uncorrelated speckle intensity patterns in sequence in a linear resist. The pdf for each exposure is eq. (7) and the pdf for the total exposure is the result of convolving the M identical pdfs.[17] The pdf for the phase ψ_s due to this total exposure is the gamma density[1]

$$p(\psi_s) = \frac{\psi_s^{M-1}}{\Gamma(M)} \left(\frac{M}{\langle \psi_s \rangle} \right)^M \exp\left[-\frac{M \psi_s}{\langle \psi_s \rangle} \right] \tag{21}$$

where $\langle \psi_s \rangle / M$ is proportional to the exposure energy of an individual speckle pattern. The effective complex amplitude is known to be the characteristic function of the pdf evaluated at frequency equal to unity[2] and thus, the complex amplitude is of the form of the M'th power of eq. (8)

$$\langle a \rangle = \left[1 + \left(\frac{\langle \psi_s \rangle}{M} \right)^2 \right]^{-M/2} \exp\left[j \left(M \operatorname{atan} \frac{\langle \psi_s \rangle}{M} \right) \right] \quad . \tag{22}$$

Eqs. (21) and (22) approximately (though with good accuracy) describe the case for time average intensity recording of speckle through a spinning ground glass diffuser (ref. 1, ch. 4.) The analysis given there interprets M as an effective number of statistically independent speckle patterns. The parameter M need not be an integer and for a fixed velocity diffuser M is proportional to exposure time.

The effective amplitudes and phases are plotted in fig. 8 against average exposure and for various values of M. The markers on the curves indicate specific points for which the effective phase shift is 2.5π. For these markers the amplitude varies between .031 and .95 for M between 10 and 602. [For M=1 the results are the same as eq. (8).] Near unity amplitudes can be produced but not for all values of phase. It may not be practical to increase M further as this increases recording time. One way to address this wide variation in M is to control multiple parameters such as intensity, diffuser angular velocity, and radial position of the laser beam on the diffuser. This would allow a modest range of control (less than 10:1) on each of the three parameters. It may also be desirable to add a separate phase bias ψ_b for amplitudes that are close to unity in order to reduce recording time.

The principal advantage of time-averaged recording is that the maximum recording depth is substantially less than for non-averaged recording. This is shown in fig. 9 for the density functions corresponding to M=2, 10 and 29 and effective amplitude a_p=.025 . The effective values of phase shift are respectively $.9\pi$, 2.6π and 4.6π . These curves can be compared with fig. 7. They are substantially narrower than the exponential density. The curves for M=1 to 10 and M=10 to 29 both produce a 2π range, however the second set of curves (compare M=2 to M=29) are even narrower. Also the exposure energy used for time-averaged recording will the be smaller by a factor of 2 to 4. This can be seen by inverting the amplitude in eq. (22) for average exposure energy

$$\langle E_s \rangle \propto \langle \psi_s \rangle = M \sqrt{a_p^{-2/M} - 1} \quad . \tag{23}$$

For a_p=.025 and M=1, which corresponds to the exponential distribution, the average intensity is proportional to 12.7π. For M=2 the exposure drops to 4.0π . For M=5 the energy is minimum at 2.9π and it increases gradually to 5.0π at M=29 . Therefore, both exposure energy and film thickness can be much less if temporal averaging is used.

Since we are only concerned about the modulo value of effective phase, there are multiple exposure conditions that produce the same value of desired amplitude $a_c=a_p$. One choice of exposure conditions may be preferable from various considerations of energy efficiency, accuracy and recording time. We illustrate this by expressing the amplitude a_p in terms of the phase ϕ_p in eq. (22). This is done by substituting out $\langle\psi_s\rangle/M$ that is in common between the expressions of amplitude and phase which gives

$$a_p = \left|\cos\frac{\phi_p}{M}\right|^M \quad . \tag{24}$$

This is plotted in fig. 10 for $\phi_p=\pi/2$, $5\pi/2$ and $9\pi/2$ as a function of M. Fig. 10 also plots $\langle\psi_s\rangle$ which is proportional to average exposure energy. The markers in these figures are identical to the circular markers in Fig. 8. A second set of diamond-shaped markers indicates alternate combinations of recording time (M) and exposure energy that produce the same effective complex amplitudes. For the smallest amplitude marker (.031) the $9\pi/2$ curve does use more energy than the $5\pi/2$ curve, but the amplitude is less sensitive to exposure time. For the two larger amplitude markers the alternate choices on the $\pi/2$ curve use less energy but are much more sensitive to exposure time. Fig. 8 shows that the sensitivity of the amplitude with respect exposure energy generally decreases with increasing exposure energy.

The gamma density function in eq. (21) can be approximated as a gaussian of the form

$$p(\psi_s) \approx \frac{1}{\sqrt{2\pi M}\langle\psi_s\rangle}\exp\left[\frac{-1}{2M}\left(\frac{\psi_s-\langle\psi_s\rangle}{\langle\psi_s\rangle}\right)^2\right] \tag{21a}$$

for M a large number, through the use of the central limit theorem. [see ref. 17 pp. 214-221, 240] Substituting this result in eq. (3) approximates the effective amplitude of eq. (22) as

$$\langle a\rangle \approx \exp\left(j\langle\psi_s\rangle\right)\exp\left(\frac{-\langle\psi_s\rangle^2}{2M}\right) \tag{22a}$$

This result is quite good for M greater than 10. Eq. (21a) is used in the next section, *Issue 7*.

Issue 7: Time averaged recording in log resist. The analysis of effective amplitude is identical to that used in deriving eq. (18) except that the gamma density is used in place of the exponential density. This gives

$$\langle a\rangle = \exp\left(j\psi_b\right)\int_0^\infty \frac{x^{M-1}}{\Gamma(M)}\exp(-x)\exp\left[j\alpha\ln\left(1+\frac{\gamma x}{M}\right)\right]dx \tag{25}$$

The amplitude again depends on the ratio of speckle to bias intensities. For M equal to 1 eq. (25) is identically eq. (18). For any value of M the amplitude decreases monotonically with increasing γ. For M a large number the gamma density in eq. (25) can be replaced by its approximate form eq. (21a). After appropriate change of variables eq. (25) is approximated as

$$\langle a \rangle \approx \frac{\exp(j\psi_b)}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-x^2/2)\, \exp\left[j\alpha \ln\left(1+\gamma+\gamma x/\sqrt{M}\right)\right] dx \qquad \textbf{(25a)}$$

Factoring out the term $1+\gamma$ in the log function and using the approximation $\ln(1+z) \approx z$ for values of z less than 1, eq. (25) further simplifies to

$$\langle a \rangle \approx \exp\left[j\psi_b + j\alpha \ln(1+\gamma)\right] \exp\left[\frac{-1}{2M}\left(\frac{\alpha\gamma}{1+\gamma}\right)^2 \right] . \qquad \textbf{(25b)}$$

The range of validity of the expansion depends on the extent of the gaussian in eq. (25a). The gaussian is essentially zero for x in excess of 3. This leads to $M > 9[\gamma/(1+\gamma)]^2$ which is always true for M greater than 9.

Eq. (25b) shows that the effective amplitude a_p monotonically decreases with increasing speckle to bias ratio γ. For low sensitivity resist (see *Issue 5*) the curves saturate without reaching zero. Increasing M only raises the saturation value, and does not increase depth of amplitude modulation over recording without time-averaging. The amplitude control provided by time-averaging in log photoresist is similar to that for time-averaged recording in linear resists as can be seen by comparing eq. (22a) to eq. (25b). The main difference between the two results is that eq. (22a) always approaches zero given a large enough exposure while eq. (25b) instead settles to a constant amplitude determined by α^2/M.

Issue 8: Spatial resolution of linear and log resists. Photoresists generally have much higher spatial resolution than the diffraction limit. However, if speckle is reimaged through a projection system it would be possible to use an adjustable iris in place of the spinning diffuser. The blurred speckle pattern can then be considered as spatially integrated. The problem is analyzed in ref. 1, ch. 2 and it is not surprising that the results are identical to the analysis of time integrated speckle given above. As above, the gamma function is a good approximation of the probability density function of the spatially averaged speckle intensities. The parameter M is now interpreted as the effective number of speckles averaged together in a rectangular window. Thus the results presented above in *Issues 6-7* can be used without modification to analyze the effect of resolution loss in linear resists.

5. Experimental demonstration of speckle recording

The theory presented in Sec. 4 primarily describes the complex amplitudes that could be produced by recording laser speckle in photoresist. In order to better anticipate the potential problems in developing the proposed exposure system we have also performed some preliminary experiments in which we use a phase-only liquid crystal light valve to represent photoresist. Unlike the demonstration reported in Sec. 2, in which the SLM represented an array of pixels, in this section the SLM represents a single pixel. The purpose of the experiment is to show the control of effective amplitude by varying the exposure.

The SLM chosen for this demonstration is a gallium arsenide photodetector, birefringent liquid crystal light valve from the Lebedev Physical Institute. It was chosen because it produces the largest phase shift (up to 4π) of the SLMs available to us. Measurements in an interferometer of the read

side of the light valve indicate that there is a roughly logarithmic dependence of the phase modulation depth on the exposure intensity. However, the exact phase shifts measured can vary dramatically based on the spatial frequency content of the illumination and the exposure intensity. In particular, the spatial resolution of the device (4 to 40 lp/mm) is known to depend on the exposure intensity. For these reasons, rather than attempting to completely characterize the device, we have chosen to demonstrate the control of effective amplitude by varying the speckle recording parameters. The experimental results for the light valve are compared with our theory for ideal logarithmic resists.

The write side of the light valve is illuminated by two mutually incoherent (850 nm) laser diode sources. One beam is expanded and illuminates the light valve with a spatially uniform bias. The other beam is focused into a small spot on the surface of a ground glass diffuser to produce a speckle pattern illumination on the light valve. The speckle diameter is varied by translating a diffuser along the path of the beam so as to change the beam diameter intercepting the diffuser. The light valve is electrically driven with a 2 kHz, 10 V rms sinusoidal potential from a signal generator. The read side of the light valve is illuminated with a 633 nm wavelength HeNe laser beam. The beam is spatially filtered and expanded using a collimator. The collimator lens is positioned to slightly converge the beam. At the face of the SLM the beam is 14.5 mm in diameter. The reflected beam is observed using a ccd camera positioned at the focus of the collimator lens. A digital oscilloscope connected to the video output of the camera is used to measure the intensity of the specular diffraction peak as the bias and speckle exposures are varied.

Experimental Results. The experimental results are plotted in fig. 11 (solid lines). For these plots the effective amplitude a_p is considered to be the square root of the intensity of each measurement after it has been normalized by the maximum intensity from all the measurements. For each set of measurements the speckle intensity is varied from 0 to 365 $\mu W/cm^2$. For one set (fig. 11a) the bias E_b is stepped from 0 to 29.3 $\mu W/cm^2$ (with the speckle diameter held constant at 1 mm.) The actual value of bias used in each experiment is given in the first column of Table 1. The top curve in fig. 11a corresponds to the first row in the table, the second curve corresponds to the second entry, and so on. For the other set (fig. 11b) the speckle diameter is stepped from .07 to 1 mm (with the bias held constant at zero.) The measured speckle diameters are listed in the first column of Table 2. The curves in fig. 11b and Table 2 are also sequenced so that the top curve corresponds to the first row, and so on.

The measured curves in fig. 11 demonstrate that effective amplitudes between 1 and nearly zero can be produced with this particular device. Both increasing bias and decreasing speckle size increase the minimum possible amplitude. The reduction in modulation range as a function of bias (fig. 11a) is probably due to the limited phase modulation range of the SLM. Increasing the bias reduces the range of phase modulation produced by the speckle portion of the exposure and thus, saturation effects (comparable to finite thickness in resists) become more prevalent. The third column of Table 1 lists the phase shifts ψ_b that are produced for the bias levels given in column 1. The phase shifts were measured using an interferometer to observe fringe displacement as a function of exposure intensity. In fig. 11b the reduction in the range of amplitude modulation with decreasing speckle diameter is apparently due to spatial averaging related to the limited resolution of the SLM. Further evidence of this is that a curve for 3 mm diameter speckle (not shown) is nearly identical to

the 1 mm curve over most of the range of γ. The only apparent discrepancy is around the minimum of each curve where the 3 mm case only dips to .16 instead of .09. . We believe that this difference is mainly due to the increased level of background noise for the 3 mm case, which is anticipated as a direct result of its lower directivity (18:1 for the 3 mm case as opposed to 165:1 for the 1 mm case.)

Comparison of experiment with theory for logarithmic resist. The effective amplitude produced by a resist film having logarithmic slope α=1.65 and different levels of bias has been calculated using eq. (19) and is plotted in fig. 11a. For low energy speckle exposures (i.e. low values of γ) the effective amplitude is unaffected by the finite thickness of the resist and thus, all theoretical curves are identical in this region. However, the SLM does not actually behave as a single function of γ for all values of bias. Instead, the experimental curves in fig. 11a have been plotted so that the initial slope of each curve matches that of the theoretical curves. This corresponds to using values of bias from the third column of Table 1 (instead of the first column) in the calculation of $\gamma=\langle E_s\rangle/E_b$. The middle two experimental curves do have the same initial slope. The value of α=1.65 has been chosen so that the initial slope of the theory and experiment match for the measured values of bias. In this theory, increasing the bias level corresponds to reducing ψ_{ms} the total phase range available for speckle recording. The values of ψ_{ms} used for the theoretical plots in fig. 11a are listed in the fourth column of Table 1. If the SLM were to closely fit the model of the log resist then we would expect that the total phase range of the resist $\psi_m = \psi_b + \psi_{ms}$ (the sum of the third and fourth columns of Table 1) would be a constant for each level of bias, rather than between 2π to 3π. While the SLM and ideal resist are significantly different, we believe that these comparisons are helpful in better anticipating the practical issues of using resists and in appreciating the theory of speckle recording.

We continue these comparisons for the recording of spatially-averaged speckle (fig. 11b.) The effective amplitude is calculated for a resist film exposed to averaged speckle. The equation used is not explicitly presented. It combines the results for thin logarithmic resists [from eq. (19)] with the results for time averaged resists [eq. (25)], and it can be derived directly by using the gamma density function for $p(\psi)$ in eq. (9). The effective amplitude for all the theoretical curves have been calculated for a single value of film thickness corresponding to $\psi_{ms}=2.26\pi$, resist sensitivity α=2.0, and the values of M listed in the third column of Table 2. These values have been selected so that the middle theoretical curve (for M=3) in fig. 11b closely fits the experimental results. With respect to the experimental curves, the other four theoretical curves appear to be overly compressed (for the lower two curves) or overly expanded (for the upper two curves) along the γ coordinate.

We also compare the numerical values of M with the measured values of diameter (Table 2.) Since M (the number of speckle images averaged together) is inversely proportional to the square of the speckle diameter, we have arbitrarily selected the proportionality constant so that the speckle diameter for experiment and theory is identical for the third row of Table 2. The other values of diameter (the second column of Table 2) are then calculated from M using this proportionality. As with the corresponding curves of fig. 11b, the difference in diameter between the theory and experiment depart increasingly as M differs from 3.

Discussion of these results. While the optical properties of the SLM and the idealized resist are quite different similar trends are apparent. As discussed in *Issue 5* of Sec. 4, low values of

sensitivity α limit the minimum achievable value of effective amplitude for a logarithmic resist; and, as discussed in *Issue 4*, finite phase modulation range ψ_{ms} causes the effective amplitude to increase for large intensity speckle exposures. In fact, the phase modulation range is so small that any level of bias exposure at all reduces the total range of effective amplitude modulation. These characteristics seem to also qualitatively describe the behavior of the SLM, which we know has low (though signal dependent) sensitivity and phase modulation range. For practical recording of arbitrary complex values we clearly need greater phase range and sensitivity; especially, since applying any bias (which is intended to realize the correct phase) further reduces the range of the effective amplitude. Likewise, speckle averaging reduces the depth of modulation, which limits our ability to achieve all complex values. These limitations reflect the shortcomings of using SLMs as demonstration vehicles, rather than of the concept of speckle recording itself. As described in *Issue 4*, there are many resists that are adequately sensitive and which can be spun on in adequately thick layers.

5. Summary and Conclusions

In this paper we have presented the concept of the patterned diffuser array in which each pixel is directly synthesized or encoded on a pixel by pixel basis. The main application of this device is the realization of complex valued spatial filters (e.g. composite pattern recognition filters, spot array generators and structured light illuminators) with phase-only structures. A second potential application of patterned diffuser arrays is as gray level photomasks for projection printing.

We have also proposed a photoresist exposure system for the custom fabrication of diffuser arrays by exposing individual pixels to appropriate combinations of spatially uniform and nonuniform illumination. The exposure system appears to place no critical requirements on optical components, vibration isolation, or air cleanliness. For this reason we believe that the components required to construct a turn key system would cost well under $100,000. The most costly component appears to be the translation stages, which should be as fast as possible to reduce fabrication time. If multiple copies of a diffuser array are required, then greater speeds are possible by using various replication methods.[16]

Patterned diffuser arrays provide a direct way to implement complex-valued modulation without resorting to numerically intensive design procedures. This approach could be used to significantly shorten the time required to design and, in many cases, to fabricate, a wide variety of diffractive optics functions.

Acknowledgments. We thank Professor Kevin M. Walsh of the University of Louisville for advice on photoresist properties and processing methodology. We also thank Hughes-JVC for the loan of the crt and optics for addressing the Hughes light valve and A. V. Parfenov of the Lebedev Physical Institute for the loan of the gallium arsenide light valve. This research was sponsored by Advanced Research Projects Agency through Rome Laboratory contract F19628-92-K0021, Army Research Office contract DAAH04-93-G-0467, and National Aeronautics and Space Administration cooperative agreement NCCW-60 through Western Kentucky University.

References

1. J.C. Dainty, Ed., *Laser Speckle and Related Phenomena*, 2nd ed. Springer-Verlag, Berlin.(1984)
2. J.W. Goodman, *Statistical Optics*, Wiley, New York. (1985)
3. R. W. Cohn and M. Liang, "Approximating fully complex spatial modulation with pseudo-random phase-only modulation," *Appl. Opt.*, **33**(20), 4406-4415. (10 July 1994)
4. R.W. Gerchberg and W.O. Saxton, "Practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik* 35, 2, 237-250 (April 1972).
5. N.C. Gallagher and B. Liu, "Method for computing kinoforms that reduces image reconstruction error," *Appl. Opt.*, 12, 10, 2328-2335 (1973).
6. F.B. McCormick, "Generation of large spot arrays from a single laser beam by multiple imaging with binary phase gratings," *Opt. Eng.*, 28,4, 299-304 (April 1989).
7. M.P. Dames, R.J. Dowling, P. McKee, and D. Wood, "Efficient optical elements to generate intensity weighted spot arrays: design and fabrication," *Appl. Opt.*, 30, 19, 2685-91 (1991).
8. E. G. Johnson, M. A. Abushagur, "Microgenetic-algorithm optimization methods applied to dielectric gratings," *JOSA A*, **12**(5), 1152-1160. (May 1995)
9. W.-H. Lee, "Computer-Generated Holograms: Techniques and Applications," Ch. 3, pp. 119-232. *Progress in Optics*, 16. E. Wolf, ed. North-Holland, Amsterdam. (1978)
10. W. J. Dallas, "Computer-Generated Holograms," Ch. 6, pp. 291-366. *The Computer in Optical Research*, B. R. Frieden, Ed. Springer, Berlin. (1980)
11. J. P. Kirk and A. L. Jones, "Phase-only Complex-Valued Spatial Filter," *JOSA*, **61**(8), 1023-1028. (August 1971)
12. T. J. Suleski and D. C. O'Shea, "Gray-scale masks for diffractive-optics fabrication: I. Commercial slide imagers," *Appl. Opt.*, **34**(32), 7507-7517. (10 November 1995)
13. D. C. O'Shea and W. S. Rockward "Gray-scale masks for diffractive-optics fabrication: II. Spatially filtered halftone screens," *Appl. Opt.*, **34**(32), 7518-7526. (10 November 1995)
14. B. Wagner, H.J. Quenzer, W. Henke, W. Hoppe, and W. Pilz, "Microfabrication of complex surface topographies using grey-tone lithography," *Sensors and Actuators A*, V.46-47, pp. 89-94. (1995)
15. T. R. Jay and M. B. Stern, "Preshaping photoresist for refractive microlens fabrication," *Opt. Eng.*, **33**(11), 3552-3555. (November 1994)
16. M. T. Gale, M. Rossi, J. Pedersen and H. Schutz, "Fabrication of continuous relief micro-optical elements by direct laser writing in photoresists," *Opt. Eng.*, **33**(11), 3556-3566. (November 1994)
17. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. McGraw-Hill, New York. (1991)
18. Shipley Corporation Microposit Products catalog.
19. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, p. 318, eq. 3.382.4. Academic Press, New York. (1980)

Table and Figure Captions

Table 1. Parameters specific to the measured and theoretical curves in fig. 11a. Also the measured speckle diameter is 1 mm, and the theory uses α=1.65 radians.

Table 2. Parameters specific to the measured and theoretical curves in fig. 11b. The measured bias E_b=0 μW/cm^2. The theory uses bias E_b=5.1 μW/cm^2, α=2.0 radians and ψ_{ms}=2.26π.

Figure 1. Controlling specular intensity by varying surface roughness.

Figure 2. An array of diffusers that produces a custom complex-valued modulation.

Figure 3. Comparison of diffraction patterns from random encoding a 100 x 100 array of desired complex values to a 100 x 100 array of phase-only pixels and to a 100 x 100 array of diffuser pixels. Each diffuser pixel is a 3x3 array of random phases that are randomly encoded to produce the same effective value of amplitude a_p. Grayscale images of diffraction pattern intensity for arrays of (a) phase-only pixels, theory; (b) diffuser pixels, theory; © phase-only pixels, experiment; and, (d) diffuser pixels, experiment. The on-axis or DC component, upper left of © and (d) is primarily due to Fresnel reflection from the cover glass which has not been anti-reflection coated for this LCLV.

Figure 4. Proximity exposure systems for producing complex-valued pixels. Phase offsets produced by (a) time averaged recording of speckle patterns from spinning diffuser, (b) adding a spatially uniform exposure which, as shown, is derived from a single mode optical fiber used as a point source.

Figure 5. Grayscale intensity images of speckle patterns recorded at (a) 0 μm, (b) 100 μm, and © 500 μm past a 100 μm slit. The diameter of the speckle is approximately 2.5 μm. Patterns were imaged onto a 1/3" ccd camera using a 40x microscope objective approximately 160 mm from the ccd. The images were then recorded with a video frame grabber.

Figure 6. Effective amplitude a_p and phase ϕ_p for log and linear resists. For linear resist, which depends on absolute intensity, the x-axis is defined to be $\gamma=\langle\psi_s\rangle/\pi$.

Figure 7. The probability density functions for depths of speckle recorded into log and linear resists. Each distribution produces effective amplitude a_p=.025 .

Figure 8. Effective complex amplitude for time-averaged recording of speckle in linear resist. Average recorded depth is proportional to average exposure energy $\langle E_s\rangle$. For M of 80, 180, 602 the effective phase curves are nearly identical. The dots (•) indicate where effective phase is 2.5π.

Figure 9. Probability density functions for time-averaged recording in linear resist. Each pdf produces identical effective amplitude a_p=.025 .

Figure 10. Effective amplitude for time-averaged recording in linear resist for a constant value of effective phase. The dots (•) indicate identical points from fig. 8. The diamonds (♦) indicate points identical in amplitude but that differ in phase by an integer multiple of 2π.

Figure 11. Experimental demonstration of speckle recording using a phase-only liquid crystal light valve to represent a photoresist. The plots show how the effective amplitude curves change for (a) different levels of uniform bias E_b and (b) different speckle diameters. Specific values used in the experiment and theory are given in Tables 1 and 2.

E_b (μW/cm^2)		ψ_b measured	ψ_{ms} theory
Measured	Theory	(radians)	(radians)
29.3	12.3	2.0π	0.91π
15.3	15.3	1.5π	1.26π
9.8	9.8	1.0π	1.67π
0.0	5.1	0.0π	2.06π

Table 1. Parameters specific to the measured and theoretical curves in fig. 11a. Also the measured speckle diameter is 1 mm, and the theory uses α=1.65 radians.

Speckle diameter (mm)		M
Measured	Theory	Theory
0.07	0.14	10.0
0.15	0.20	4.5
0.25	0.25	3.0
0.4	0.35	1.5
1.0	0.43	1.0

Table 2. Parameters specific to the measured and theoretical curves in fig. 11b. The measured bias E_b=0 μW/cm^2. The theory uses bias E_b=5.1 μW/cm^2, α=2.0 radians and ψ_{ms}=2.26π.

[JOSA A Cohn, Vasiliev, Liu, Hill]

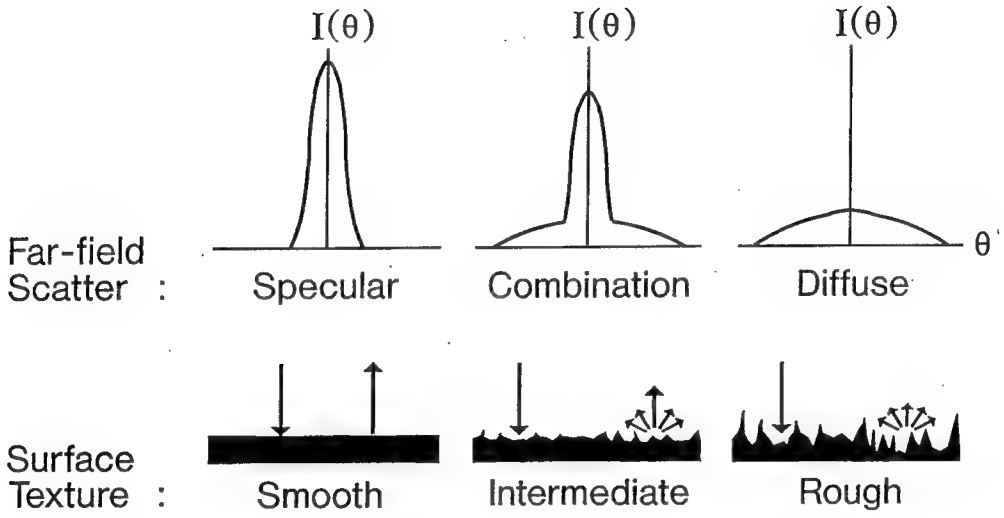


Fig. 1. Controlling specular intensity by varying surface roughness

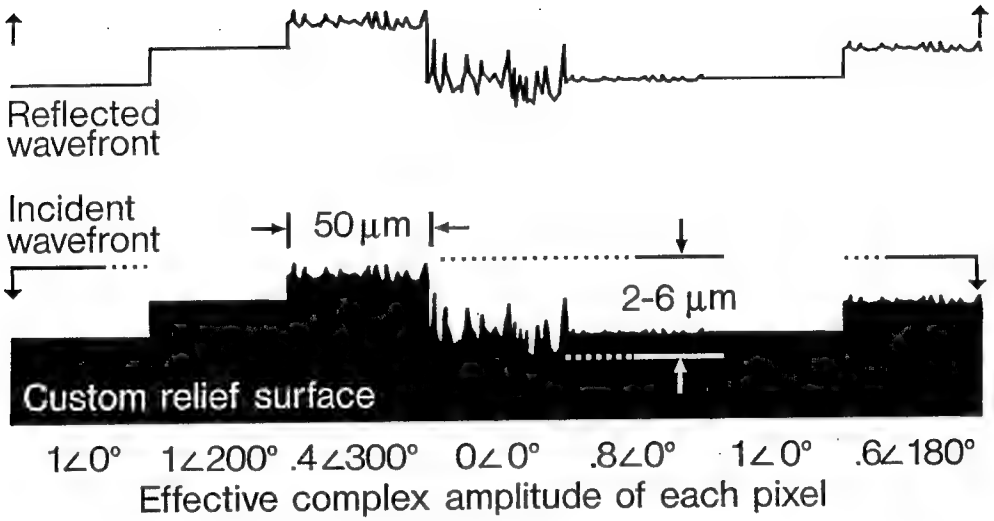


Fig. 2. An array of diffusers that produces a custom complex-valued modulation.

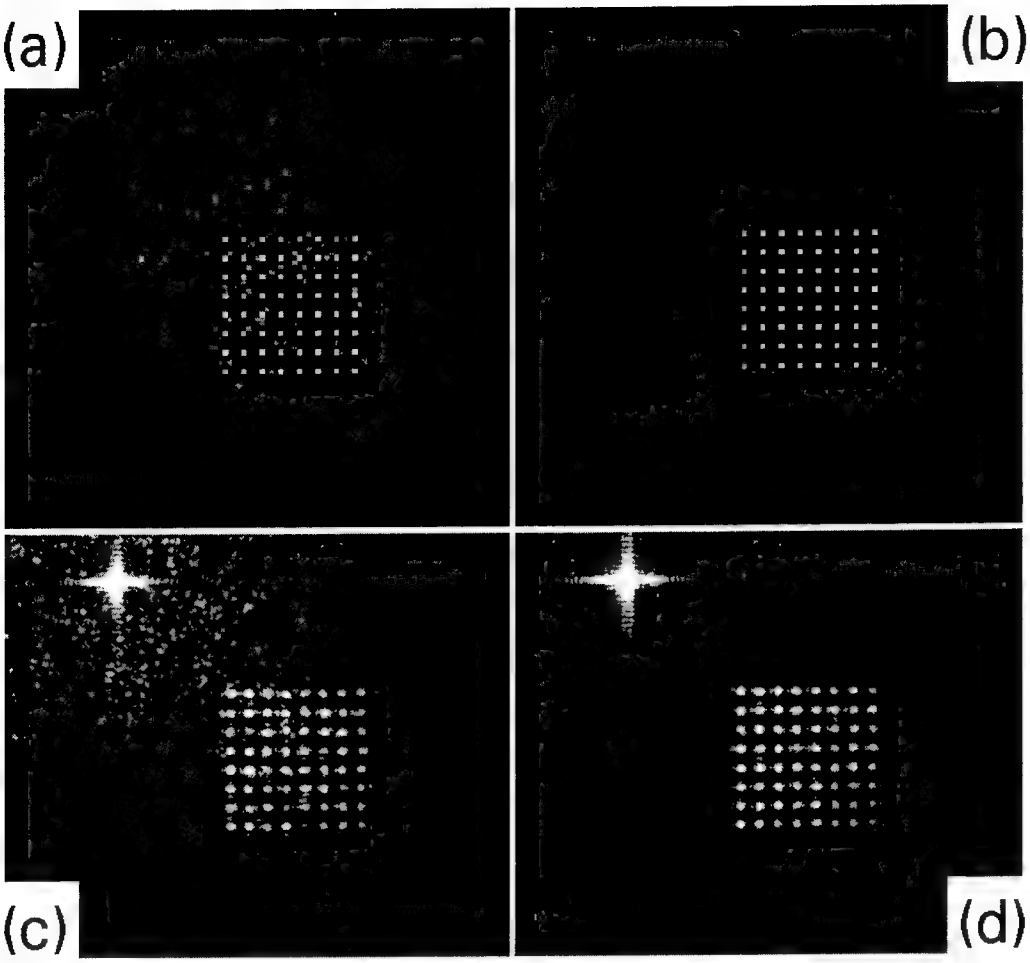


Figure 3. Comparison of diffraction patterns from random encoding a 100×100 array of desired complex values to a 100×100 array of phase-only pixels and to a 100×100 array of diffuser pixels. Each diffuser pixel is a 3×3 array of random phases that are randomly encoded to produce the same effective value of complex amplitude a_p. Grayscale images of diffraction pattern intensity for arrays of (a) phase-only pixels, theory; (b) diffuser pixels, theory; (c) phase-only pixels, experiment; and, (d) diffuser pixels, experiment. The on-axis or DC component, upper left of (c) and (d) is primarily due to Fresnel reflection from the cover glass which has not been anti-reflection coated for this LCLV.

[JOSA A Cohn, Vasiliev, Liu, Hill]

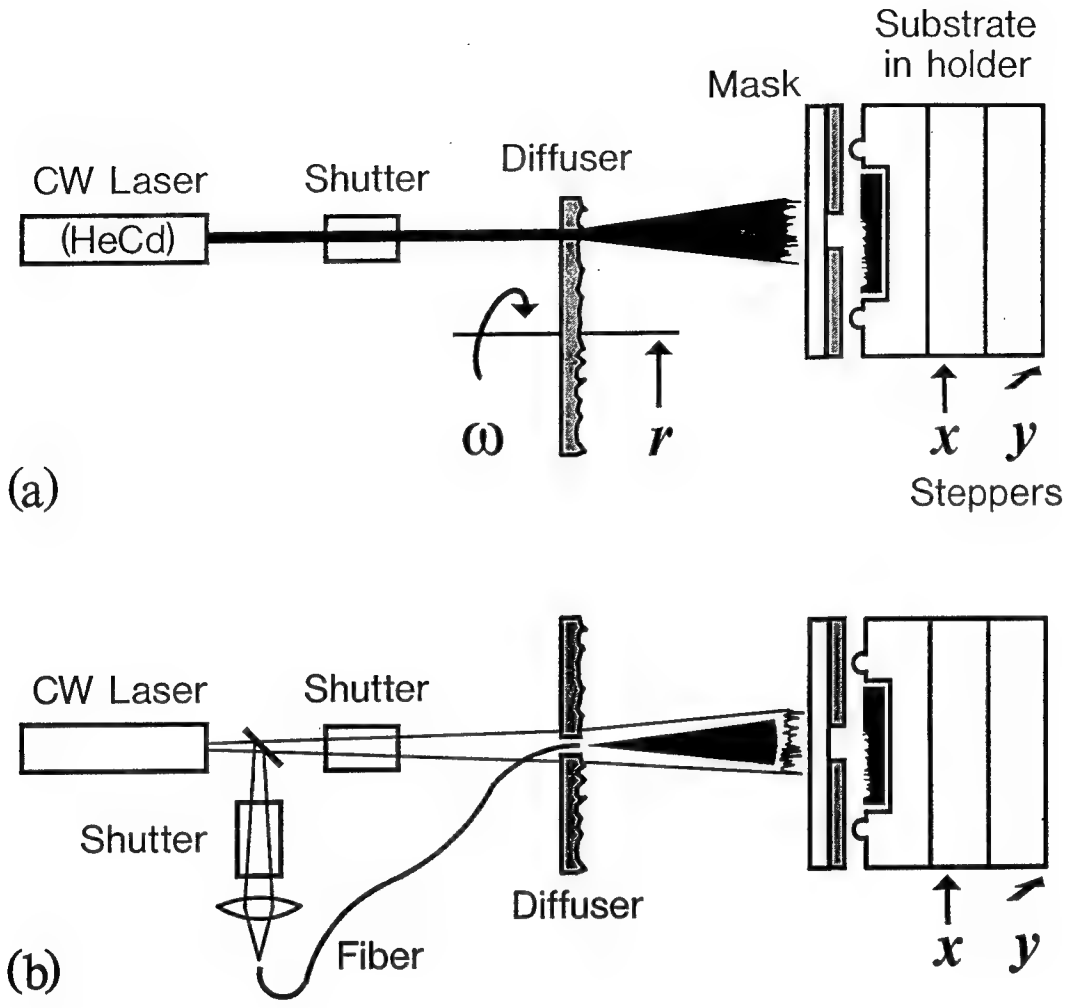


Fig. 4. Proximity exposure systems for producing complex-valued pixels. Phase offsets produced by (a) time averaged recording of speckle patterns from spinning diffuser, (b) adding a spatially uniform exposure which, as shown, is derived from a single mode optical fiber used as a point source.

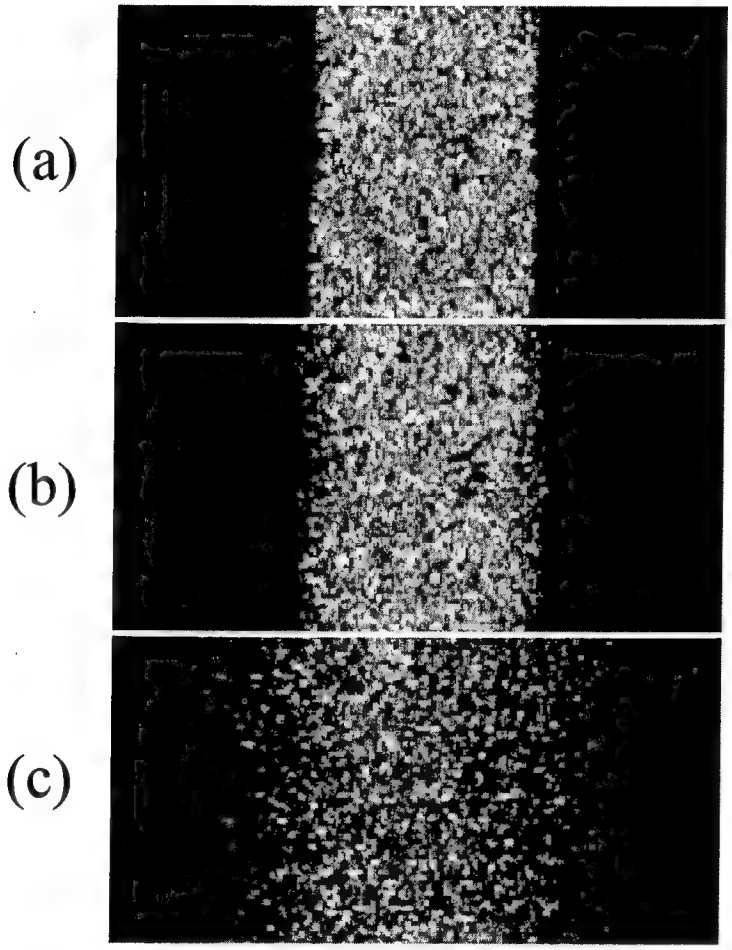


Figure 5. Grayscale intensity images of speckle patterns recorded at (a) 0 μm, (b) 100 μm, and (c) 500 μm past a 100 μm slit. Patterns were imaged onto a 1/3" ccd camera using a 40× microscope objective approximately 160 mm from the ccd. The video images were then recorded with a video frame grabber.

[JOSA A Cohn, Vasiliev, Liu, Hill]

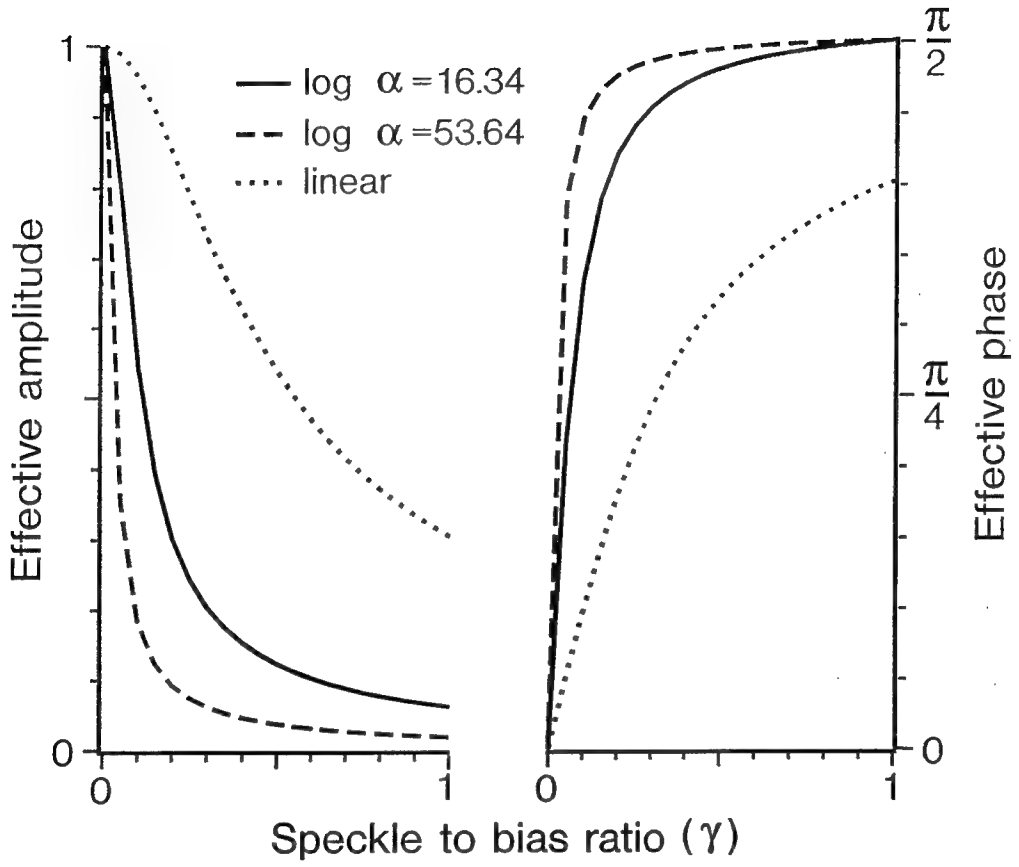


Fig. 6. Effective amplitude a_s and phase ϕ_s for log and linear resists. For linear resists, which only depend on absolute intensity, the x-axis is defined to be $\gamma = \langle \Psi_s \rangle / \pi$.

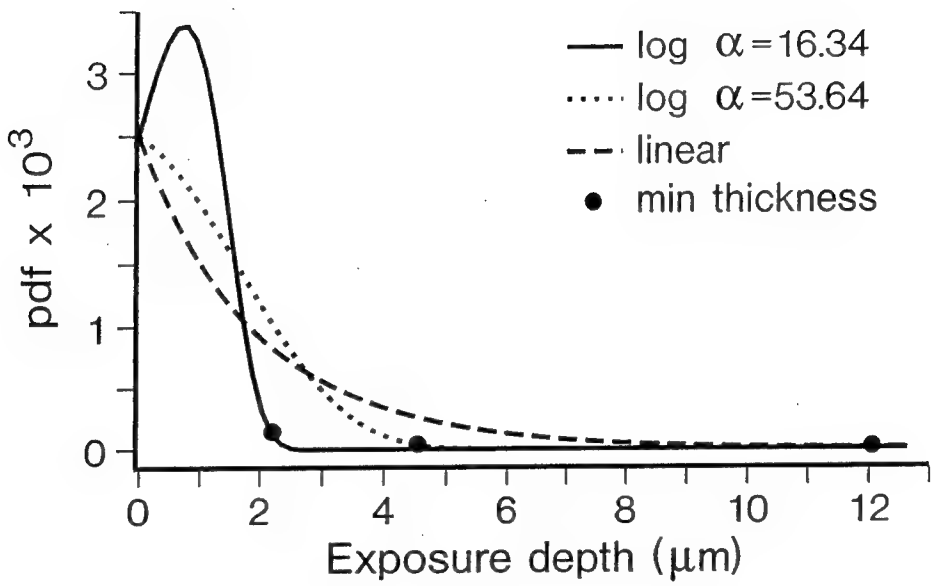


Fig. 7. The probability density functions for depths of speckle recorded into log and linear resists. Each distribution produces effective amplitude a_e=.025 .

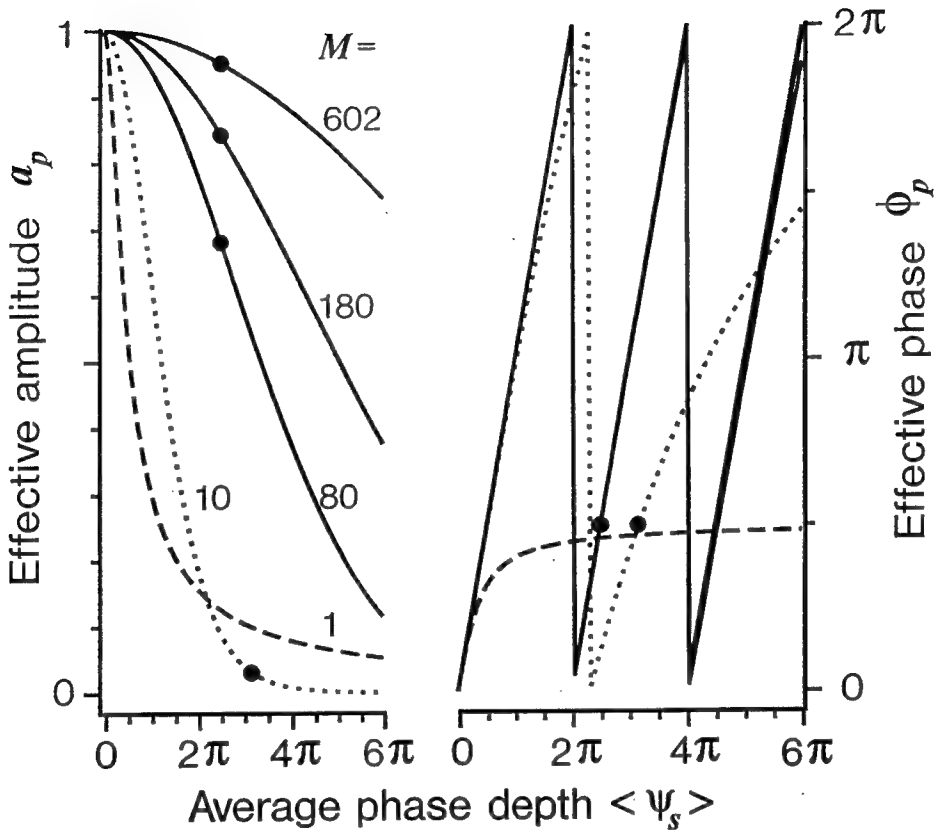


Fig. 8. Effective complex amplitude for time-averaged recording of speckle in linear resist. Average recorded depth is proportional to average exposure energy $<E>$. For M of 80, 180 and 602 the effective phase curves are nearly identical. The dots (•) indicate where effective phase is 2.5π.

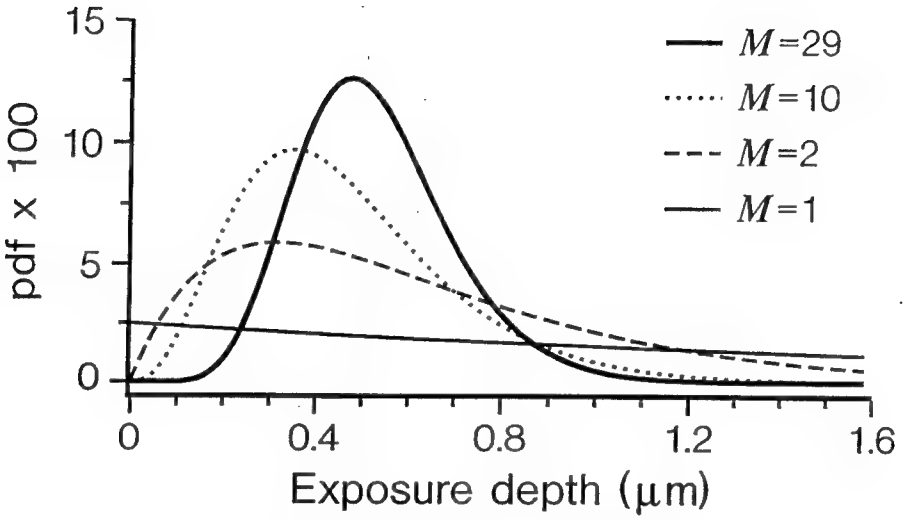


Fig. 9. Density functions for time-averaged recording in linear resist. Each pdf produces identical effective amplitude of a_e=.025 .

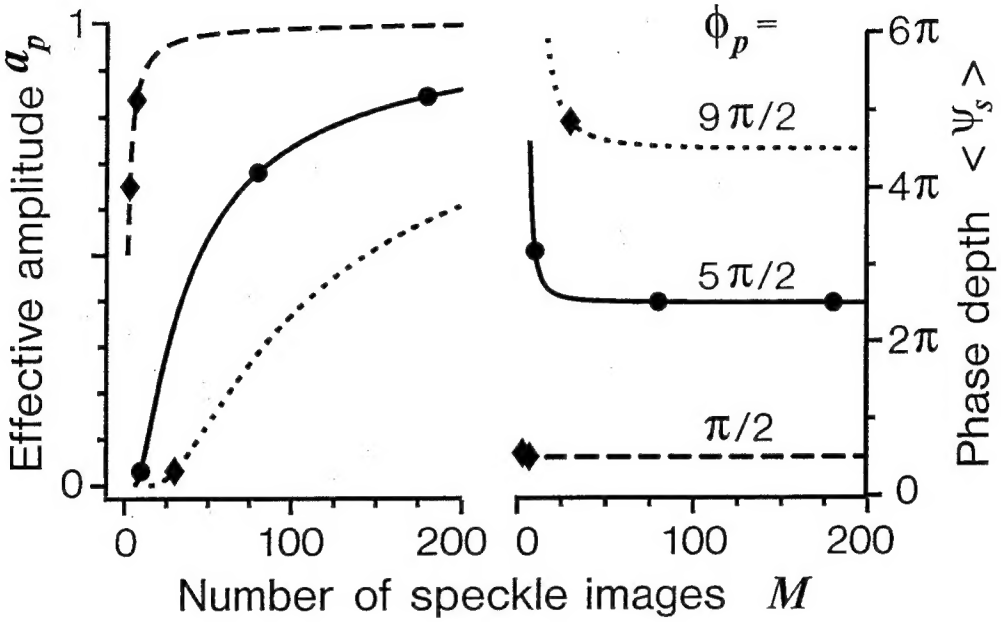


Fig. 10. Effective amplitude for time-averaged recording in linear resist for a constant value of effective phase. The dots (●) indicate identical points from fig. 8. The diamonds (♦) indicate points identical in amplitude but that differ in phase by an integer multiple of 2π .

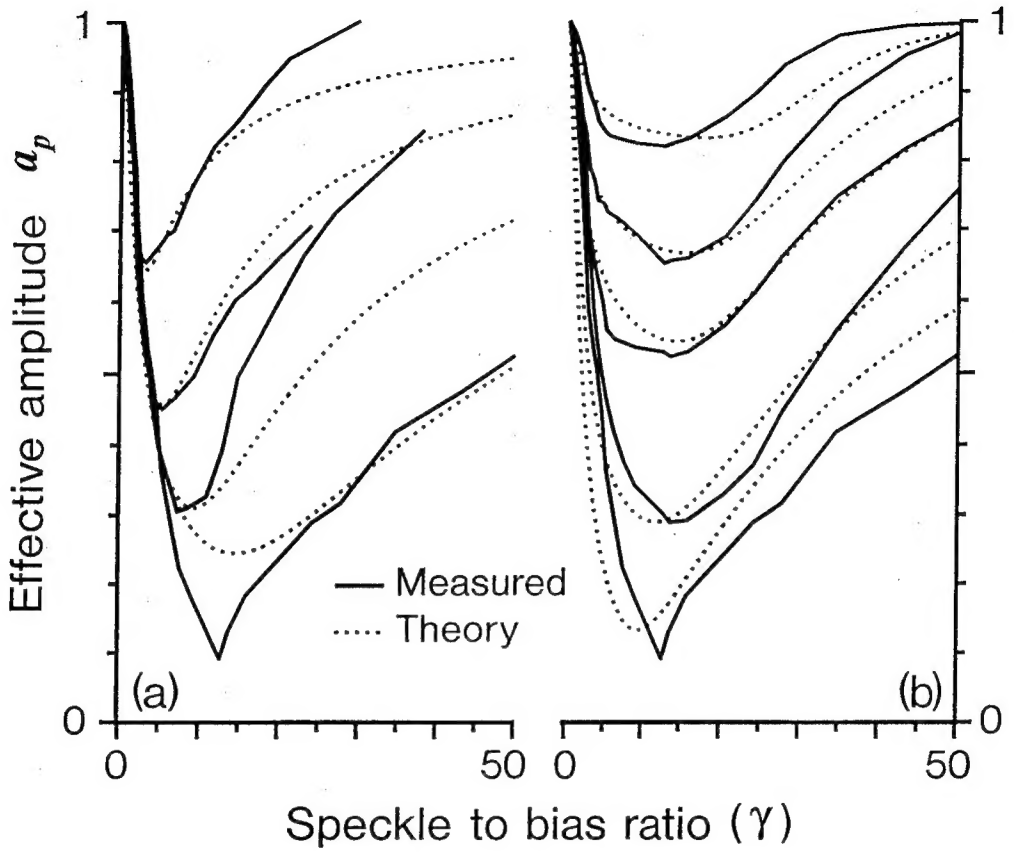


Fig. 11. Experimental demonstration of speckle recording using a phase-only liquid crystal light valve to represent a photoresist. The plots show how the effective amplitude curves change for (a) different levels of uniform bias E_b and (b) different speckle diameters. Specific values used in the experiment and theory are given in Tables 1 and 2.

*U.S. GOVERNMENT PRINTING OFFICE: 1996-710-126-47038

DISTRIBUTION LIST

addresses	number of copies
MAJOR THOMAS J. GRYCEWICZ ROME LABORATORY/EROP 80 SCOTT DRIVE HANSCOM AFB, MA 01731-2909	5
ROBERT W. COHN DEPT OF ELECTRICAL ENGINEERING UNIVERSITY OF LOUISVILLE LOUISVILLE, KY 40292	5
ROME LABORATORY/SUL TECHNICAL LIBRARY 26 ELECTRONIC PKY ROME NY 13441-4514	1
ATTENTION: DTIC-OCC DEFENSE TECHNICAL INFO CENTER 8725 JOHN J. KINGMAN ROAD, STE 0944 FT. BELVOIR, VA 22060-6218	2
ADVANCED RESEARCH PROJECTS AGENCY 3701 NORTH FAIRFAX DRIVE ARLINGTON VA 22203-1714	1
MR. BRIAN HENDRICKSON ARPA/MTO 3701 N. FAIRFAX DR ARLINGTON, VA 22203	5
MR. RICHARD JUDAY NASA/JSC, EE6 HOUSTON, TX 77058	2
MR. WILLIAM FRIDAY U. S. ARMY MISSILE COMMAND AMSMI-RD-WS-PO REDSTONE ARSENAL, AL 35898-5248	1

MISSION

OF

ROME LABORATORY

Mission. The mission of Rome Laboratory is to advance the science and technologies of command, control, communications and intelligence and to transition them into systems to meet customer needs. To achieve this, Rome Lab:

a. Conducts vigorous research, development and test programs in all applicable technologies;

b. Transitions technology to current and future systems to improve operational capability, readiness, and supportability;

c. Provides a full range of technical support to Air Force Materiel Command product centers and other Air Force organizations;

d. Promotes transfer of technology to the private sector;

e. Maintains leading edge technological expertise in the areas of surveillance, communications, command and control, intelligence, reliability science, electro-magnetic technology, photonics, signal processing, and computational science.

The thrust areas of technical competence include: Surveillance, Communications, Command and Control, Intelligence, Signal Processing, Computer Science and Technology, Electromagnetic Technology, Photonics and Reliability Sciences.